

# New Patent Application

entitled

## Chimeric Transcription Factors

by

Sridaran Natesan  
Michael Gilman

Express Mail Label # EI718326523US

Date of mailing: September 28, 1999

I hereby certify that this document is being deposited with the United States Postal Service "Express Mail Post Office To Addressee" under 37 CFR §1.10 on the date indicated above and is addressed to: Box Patent Application, Assistant Commissioner of Patents, Washington, DC 20231.

Signed: \_\_\_\_\_

*Carolyn Stewart*

## Chimeric Transcription Factors

### Table of Contents

Related Applications .....	1
Introduction .....	1
Summary of the invention .....	1
Brief Description of the Figures .....	7
Detailed Description of the Invention .....	8
Definitions .....	8
Activate .....	8
"Capable of selectively hybridizing .....	8
Cells", "host cells" and "genetically engineered cells .....	8
Cell line .....	8
Composite", "fusion", "chimeric" and "recombinant .....	9
coding sequence .....	9
construct .....	9
Derived from .....	9
Divalent .....	10
Domain .....	10
DNA binding domain .....	10
DNA recognition sequence .....	10
Endogenous .....	10
Exogenous .....	10
Gene .....	10
Genetically engineered cells .....	10
Heterologous .....	10
Interact .....	11
Ligand .....	11
Minimal promoter .....	11
"promoter" and "transcription control sequence .....	11
Nucleic acid .....	12
Oligomerization .....	12
multimerization .....	12
Operably linked .....	12
"Protein", "polypeptide" and "peptide" .....	13
"recombinant virus" .....	13
"target gene" .....	13
"Transcription factor" .....	13
"Transfection" .....	13
Infection .....	13
Transduction .....	13
Transgene .....	13
"Transient transfection .....	14
Components of the system and additional details .....	14
1. Transcription Activation Domains .....	14
2. Ligand binding domains .....	17
Table 1: .....	19
3. Ligands of the invention .....	22
4. DNA-binding domains .....	23

5. Regulatory domains .....	24
6. Additional domains .....	25
7. Promoters .....	26
8. Transcription factors, additional comments. ....	27
9. Target gene constructs .....	27
10. Design and assembly of the DNA constructs .....	33
11. Cells .....	33
12. Introduction of Constructs into Animals .....	35
13. Applications .....	37
<b>Examples</b> .....	41
I. Individual DNA-binding and transcription activating components are modular, may be incorporated into fusion proteins with various other domains and function as intended in cell culture and in animals: .....	41
A. ZFHD1 and p65 work well individually in cell culture and in whole animals in drugdependent (regulatable) transcription systems .....	41
B. Hybrid transcription factors containing such modular components work well in constitutive expression .....	60
II. Evaluation of representative illustrative chimeric transcription factors .....	63
III. Illustrative chimeric transcription factor for allostery-based systems .....	68
p65/GAL4/PR-LBD .....	68
ZFHD1/p65/PR-LBD .....	69
rtTA/p65 .....	69
IV. Regulated transcription mediated by the chimeric transcription factor S3H .....	69
<b>Claims:</b> .....	71

## Chimeric Transcription Factors

### Related Applications

This application is a continuation in part of USSN 09/262,721 filed March 4, 1999, which is a continuation in part of USSN 09/096,732 filed June 11, 1998, which is a continuation-in-part of USSN 08/672,213, filed June 27, 1996 which was a continuation of USSN 60/000,553, filed June 27, 1995 and USSN 60/019,614, filed December 29, 1995, the full contents of which are incorporated herein by reference.

### 15 Introduction

A variety of applications involving gene transcription, including among others, gene therapy, production of biological materials and biological research, depend on the ability to elicit specific and high-level expression of genes encoding RNAs or proteins of therapeutic, commercial or experimental value. Achieving a sufficiently high level of expression for clinical or other utility in genetically engineered cells within whole organisms has often been a limiting problem. Various approaches for addressing this problem, including the search for stronger transcriptional promoters or higher transfection efficiencies, have in many cases not met with success. Meanwhile, in various lines of research with transcription factors, promising results in transient transfection models have not been borne out with chromosomally integrated reporter gene constructs. Furthermore, overexpression of transcription factors is commonly associated with toxicity to the host cell. Despite those precedents, this invention takes a novel approach to the challenge of optimizing heterologous gene expression through new uses of, and new designs for, transcription factor proteins which are expressed within the engineered cells containing the target gene. The invention provides improved methods and materials for achieving high-level expression of a target gene in genetically engineered cells, including genetically engineered cells within whole organisms.

### Summary of the invention

This invention involves improvements in transcription activation domains and their use in fusion proteins referred to as "chimeric transcription factors". The invention further involves DNA sequences encoding those chimeric transcription factors, transcription control sequences responsive to the chimeric transcription factors, target gene constructs containing a target gene operably linked to such a transcription control sequence, genetically engineered cells which contain a target gene construct and a construct for expressing a chimeric transcription factor of the invention, organisms containing such cells, methods for producing



such cells and organisms, and methods for using the foregoing in gene therapy, production of biological materials and biological research.

Of particular interest are recombinant nucleic acids (typically DNA, although embodiments involving RNA are within the scope of the invention) encoding a chimeric transcription factor which (a) contains at least two mutually heterologous domains comprising a p65 domain and a ligand binding domain, and (b) is capable of activating the transcription of an appropriate target gene construct, discussed below, in a ligand-dependent manner. The p65 domain comprises part or all of the peptide sequence spanning positions 281-550, 321-550, 361 through 550, 361 through 450 or 450 through 550 of NF- $\kappa$ B p65 (preferably human), or a peptide sequence derived therefrom. A "p65 domain" as that term is used herein denotes a peptide sequence from about 40 amino acids up to several hundred amino acids in length, as discussed in further detail and illustrated below. To avoid confusion, peptide sequences derived from p65 which contain fewer than 40 amino acids are referred to as "p65 motifs". In addition to the p65 and ligand binding domains, the chimeric transcription factors may further contain one or more additional, optional domains, including for instance, one or more transcription potentiating domains and/or a nuclear localization sequence.

Preferred chimeric transcription factors stimulate transcription of a target gene in a ligand-dependant manner as disclosed in greater detail below. Preferably the difference in level of transcription observed in the presence and absence of ligand, respectively, is at least two, more preferably three, and even more preferably four or more orders of magnitude.

In various embodiments the chimeric transcription factor contains one or more copies of one or more p65 domains, optionally together with one or more copies of one or more different transcription activation domains, subdomains or potentiating motifs derived, for example, from p65 or other transcription activation domains (collectively, "transcription potentiating domains"). Transcription activation domains comprising a non-naturally occurring peptide sequence containing either two or more heterologous activation domains, one activation domain and two or more copies of a reiterated peptide sequence or two or more copies of a reiterated peptide sequence constitute "composite transcription activation domains". One illustrative class of composite transcription activation domains comprise domains containing (a) two or more p65 domains (which may be the same or different), (b) one or more p65 domains plus one or more p65 motifs, or (c) one or more p65 domains together with one or more copies of one or more transcription activation potentiating domains.

Transcription potentiating domains are peptide sequences which can be shown to potentiate the transcription activation potency of a transcription factor (relative to the corresponding chimeric transcription factor lacking that potentiating domain). Illustrative potentiating domains comprise motifs which may be selected or derived from the so-called "proline-rich", "glutamine-rich" and "acidic" activation motifs such as the VP16 V8 motif

(DFDLMLG), the related "V9" motif (DFDLMLGG), a human activation motif such as the 14 amino acid acidic motif of human heat shock factor (HSF) or an alanine/proline-rich motif selected from p53 or CTF (preferably human) including such motifs which are homologous to alanine/proline rich motifs within residues 361 - 450 of p65. Alternatively, an activation domain such as amino acids 431-529 of HSF may be fused to p65 to form a composite activation domain. An exemplary composite activation domain (referred to as S3H) comprises the peptide sequence p65(281-551) linked to the peptide sequence HSF(406-530).

Transcription factors of this invention may be operatively linked to any promoter selected by the practitioner, including strong promoters such as CMV or weaker promoters like RSV, the MCK enhancer, or promoters for endogenous human proteins.

A wide variety of ligand binding domains may be used in this invention, although ligand binding domains which bind to a cell permeant ligand are generally preferred. It is also preferred that the ligand have a molecular weight under about 5kD, more preferably below 2.5 kD and optimally below about 1500 D. Non-proteinaceous ligands are also preferred. Ligand binding domains include, for example, domains selected or derived from (a) an immunophilin (e.g. FKBP 12), cyclophilin or FRAP domain; (b) a hormone receptor such as a receptor for progesterone, ecdysone or another steroid; and (c) an antibiotic receptor such as a tetR domain for binding to tetracycline, doxycycline or other analogs or mimics thereof.

A tetR domain useful in the practice of this invention may comprise a naturally occurring peptide sequence of a tetR of any of the various classes (e.g. class A, B, C, D or E) (in which case the absence of the ligand stimulates target gene transcription), or more preferably, comprises a mutated tetR which comprises at least one amino acid substitution, addition or deletion compared to a wild-type tetR, especially those mutated tetR domains in which the presence of the ligand stimulates target gene transcription in a cell engineered in accordance with this invention. For example, mutated tetR domains include mutated Tn10-derived tetR domains having an amino acid substitution at one or more of amino acid positions 71, 95, 101 and 102. By way of further illustration, one mutated tetR comprises amino acids 1 - 207 of the Tn10 tetR in which glutamic acid 71 is changed to lysine, aspartic acid 95 is changed to asparagine, leucine 101 is changed to serine and glycine 102 is changed to aspartic acid. Ligands include tetracycline and a wide variety of analogs and mimics of tetracycline, including for example, anhydrotetracycline and doxycycline. Target gene constructs in these embodiments contain a target gene operably linked to a transcription control sequence including one or more copies of a DNA sequence recognized by the tetR of interest, including for example, an upstream activator sequence for the appropriate tet operator. See e.g. US Patent No. 5,654,168, the full contents of which are expressly incorporated by reference.

A wide variety of DNA binding domains may be used in the practice of this invention, including a domain selected or derived from a GAL4, *lexA* or composite (e.g. ZFHD1) DNA binding domain, or a DNA binding domain, e.g., in combination with ligand binding domains such as a wt or mutated progesterone receptor domain. TetR domains are discussed in the context of ligand binding domains. In many applications it is preferable to use a DNA binding domain which is heterologous to the cells to be engineered. Heterologous DNA binding domains include those which occur naturally in cell types other than the cells to be engineered as well as composite DNA binding domains containing component portions which are not found in the same continuous polypeptide or gene in nature, at least not in the same order or orientation or with the same spacing present in the composite domain. In the case of composite DNA binding domains, component peptide portions which are endogenous to the cells or organism to be engineered are generally preferred.

In the case of the chimeric transcription factors containing a tetR domain, the DNA binding domain is typically provided by the tetR component, and is by its nature heterologous to eukaryotic cells. TetR domains are discussed in further detail in the context of ligand binding domains.

In embodiments in which an endogenous gene is to be regulatably expressed, a composite DNA binding domain which is selected for recognition of one or more sequences upstream of the target gene may be deployed.

Additional information concerning DNA binding domains is provided below.

Typically it will be preferred that, to the extent available, each of the various domains incorporated into the design of the chimeric transcription factor contain a peptide sequence which is or is derived from a peptide sequence which naturally occurs in the cells or organism in which the chimeric transcription factor is to be expressed. Thus, human sequences or sequences derived therefrom are preferred for chimeric transcription factors for use in humans. In some cases, such as the ecdysone or tetR cases, the inclusion of non-human sequence may be unavoidable. The encoded chimeric transcription factor may further contain one or more additional domains such as a transcription potentiating domain.

Specific examples of such chimeric transcription factors include fusion proteins containing at least:

(a) a p65 domain and a ligand binding domain containing a peptide sequence selected from within, or derived from, an FKBP, FRB or cyclophilin domain (these will typically be paired with a second fusion protein comprising a DNA binding domain and at least one ligand binding domain which, in the presence of a divalent ligand, forms a ligand-dependent (cross-linked) complex with the p65 fusion protein and activates transcription of a target gene operably linked to a transcription control

sequence containing one or more recognition sequences for the DNA binding domain);

5 (b) a p65 domain, a DNA binding domain (e.g. GAL4 or ZFHD1) and a ligand binding domain containing peptide sequence selected from within, or derived from, a hormone receptor such as a progesterone receptor domain (see e.g. WO 93/23431 and WO 98/18925) or ecdysone receptor (see e.g., WO 97/38117 and WO 96/37609); and,

10 (c) a p65 domain and a tetR domain, or domain derived therefrom, which binds to a characteristic DNA sequence in a ligand-dependent manner (see e.g., US Patent No. 5,650,298 and 5,654,168).

The chimeric transcription factors may further include one or more optional domains, including for example, one or more transcription potentiating domains, and the p65 domain  
15 may in fact be replaced with a p65-containing composite transcription activation domain as described above.

The recombinant nucleic acid encoding the chimeric transcription factor may be operably linked to a transcription control sequence permitting expression of the chimeric transcription factor in cells. Such recombinant nucleic acid constructs may be contained within  
20 any of a variety of DNA vectors for use in transfecting prokaryotic or eukaryotic cells. A target gene construct may be included in the same vector or may be provided in an additional vector. The recombinant nucleic acid encoding the chimeric transcription factor and optionally a target gene construct may be present within one or more recombinant viruses for delivery (by infection) to cells in vitro or in vivo (i.e., by administration of recombinant virus to the  
25 whole organism). Conventional techniques may be used to prepare recombinant viruses harboring the recombinant nucleic acids of this invention. Adenoviruses, adeno-associated viruses, hybrid adeno-AAV, retroviruses and lentiviruses are of particular interest at present.

Compositions containing a recombinant nucleic acid encoding a chimeric transcription factor together with a target gene construct may be included in a kit or package for delivery to  
30 researchers, hospitals, physicians or veterinarians. In some cases a "universal" target gene construct may be included in which the target gene is replaced with a cloning site for insertion by the practitioner of any desired coding sequence. Such compositions or kits which are designed for regulated expression may further include a sample of ligand for activating target gene transcription. It should also be noted that the various nucleic acids may be present in  
35 vectors, recombinant viruses, etc. as described elsewhere.

A recombinant nucleic acid encoding a chimeric transcription factor of this invention may be used to transduce a cell to render it capable of expressing a target gene in a ligand-

dependent manner. The chimeric transcription factor is chosen which is capable of stimulating, in a ligand-dependent manner, the transcription of a target gene operably linked to a transcription control sequence recognized by the chimeric transcription factor. A target gene construct comprising a desired target gene operably linked to a transcription control sequence which is recognized by the chimeric transcription factor may be transduced into the cell as well, and may be included in the same or a different vector or recombinant virus for this purpose (as the recombinant nucleic acid encoding the chimeric transcription factor).

In certain applications, cells are so transduced in vitro. In other applications cells are transduced while present within an organism, generally a human or non-human mammal.

Cells containing a recombinant nucleic acid encoding a chimeric transcription factor of this invention are useful in a variety of applications as mentioned above, especially cells which further comprise a target gene operably linked to a transcription control sequence which is responsive to the chimeric transcription factor in the presence of a ligand. In order to stimulate transcription of the target gene in such cells, one exposes the cells to a ligand which binds to the chimeric transcription factor. This may be conveniently effected by simply adding the ligand to the culture medium, in an effective amount to yield the desired level of transcription.

Examples of such cells include the following:

(1) A cell containing (a) a recombinant nucleic acid encoding a chimeric transcription factor which comprises a p65 domain, a DNA binding domain and a ligand binding domain comprising or derived from a progesterone receptor domain, and (b) a target gene construct which comprises a target gene operably linked to a transcription control sequence which contains one or more copies of a DNA sequence recognized by the DNA binding domain of the chimeric transcription factor, the cell being capable of expressing its target gene in a ligand-dependent manner, the ligand being progesterone or an analog or mimic thereof.

(2) A cell containing (a) a recombinant nucleic acid encoding a chimeric transcription factor which comprises a p65 domain and a tetR domain which binds to a recognized DNA sequence in the presence of its ligand, and (b) a target gene construct which comprises a target gene operably linked to a transcription control sequence which contains one or more copies of a DNA sequence recognized by the tetR domain of the chimeric transcription factor, the cell being capable of expressing its target gene in a ligand-dependent manner, the ligand being tetracycline, doxycycline or an analog or mimic thereof.

(3) A cell containing (a) a recombinant nucleic acid encoding a chimeric transcription factor which comprises a p65 domain and an ecdysone receptor domain capable of binding to a DNA binding protein comprising or derived from the peptide sequence of an RXR protein, and

(b) a target gene construct which comprises a target gene operably linked to a transcription control sequence which contains one or more copies of a DNA sequence recognized by the RXR, the cell being capable of expressing its target gene in a ligand-dependent manner, the ligand being ecdysone or an analog or mimic thereof.

5

Cells so engineered to express a chimeric transcription factor of this invention and a corresponding target gene construct responsive to such factor in a ligand-dependent manner may be introduced into the host organism, thereby rendering the organism capable of regulated expression of a target gene. A non-human organism containing one or more such  
10 cells can be used in research to study the effect of regulated expression of a target gene of possible interest. Such animals may also be used as model systems for the study of various diseases and for the evaluation of drug candidates for treating such diseases.

Alternatively, the various recombinant nucleic acids may be introduced directly into the organism to transduce cells *in vivo* and render a host organism capable of regulated  
15 expression of a target gene. Of the various methods for gene delivery, use of one or more recombinant viruses containing the recombinant nucleic acids is currently preferred. Particularly important applications of this methodology involve the use of human subjects as the host organism. To stimulate transcription of a target gene in an organism containing cells transduced with the appropriate recombinant nucleic acids, one administers to the organism,  
20 by any acceptable means of administration, a ligand which binds to the chimeric transcription factor expressed in the cells, in an amount effective to yield the desired level of gene transcription. The ligand may be in the form of a pharmaceutically or veterinarily acceptable composition, delivered by any pharmaceutically or veterinarily acceptable route of administration.

25

### **Brief Description of the Figures**

Figure 1 demonstrates that *in vivo* administration of a ligand to animals into which engineered cells had been transplanted led to regulated gene expression and the production and  
30 secretion of the gene product. HT1080 cells were transfected with DNA constructs encoding regulatable transcription factor components as described in the examples below. Transfected HT1080 cells ( $2 \times 10^6$  total per animal, in four different sites) were injected intramuscularly into male *nu/nu* mice. Approximately one hour later, animals received the indicated concentration of intravenous rapamycin. Blood samples were collected 17 hours after  
35 rapamycin administration and assayed for hGH concentration. Rapamycin treatment produced a dose-dependent increase in serum hGH ( $X \pm SEM$ ;  $n =$  at least 5 at each dose). \*

represent statistical significance from each lower rapamycin dose and † represents statistical significance from rapamycin doses which are 10-fold and more lower ( $p < 0.05$ , one-way analysis of variance and Tukey-Kramer multiple comparison testing).

- 5 Figs 2 through 7 present comparative data on a representative collection of chimeric transcription factors assayed in cell lines into which target gene constructs (SEAP) had been stably integrated as described in the examples which follow below.

Figure 8 shows the potency of the transcription factor S3H (p65(281-551)-HSF(406-530)) in  
10 an in vitro transcription assay.

## Detailed Description of the Invention

### 15 Definitions

For convenience, the intended meaning of certain terms and phrases used herein are provided below.

"**Activate**" as applied to the expression or transcription of a gene denotes a directly or indirectly observable increase in the production of a gene product, e.g., an RNA or  
20 polypeptide encoded by the gene.

"**Capable of selectively hybridizing**" as that phrase is used herein means that two DNA molecules are susceptible to hybridization with one another, despite the presence of other DNA molecules, under hybridization conditions which can be chosen or readily determined empirically by the practitioner of ordinary skill in this art. Such treatments include  
25 conditions of high stringency such as washing extensively with buffers containing 0.2 to 6 x SSC, and/or containing 0.1% to 1% SDS, at temperatures ranging from room temperature to 65-75°C. See for example F.M. Ausubel et al., Eds, Short Protocols in Molecular Biology, Units 6.3 and 6.4 (John Wiley and Sons, New York, 3d Edition, 1995).

"**Cells**", "**host cells**" and "**genetically engineered cells**" refer not only to the  
30 particular subject cells but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

"**Cell line**" refers to a population of cells capable of continuous or prolonged growth  
35 and division in vitro. Often, cell lines are clonal populations derived from a single progenitor cell. It is further known in the art that spontaneous or induced changes can occur in karyotype during storage or transfer of such clonal populations. Therefore, cells derived from the cell line

referred to may not be precisely identical to the ancestral cells or cultures, and the cell line referred to includes such variants.

“**Composite**”, “**fusion**”, “**chimeric**” and “**recombinant**” denote a material such as a nucleic acid, nucleic acid sequence or polypeptide which contains at least two constituent portions which are mutually heterologous in the sense that they are not otherwise found directly (covalently) linked in nature, e.g. are not found in the same continuous polypeptide or gene in nature, at least not in the same order or orientation or with the same spacing present in the composite, fusion or recombinant product. Such materials contain components derived from at least two different proteins or genes or from at least two non-adjacent portions of the same protein or gene. In general, “composite” refers to portions of different proteins or nucleic acids which are joined together to form a single functional unit, while “fusion” generally refers to two or more functional units which are linked together. “Recombinant” is generally used in the context of nucleic acids or nucleic acid sequences.

A “**coding sequence**” or a sequence which “encodes” a particular polypeptide or RNA, is a nucleic acid sequence which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide in vitro or in vivo when placed under the control of an appropriate expression control sequence. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, cDNA from procaryotic or eukaryotic mRNA, genomic DNA sequences from procaryotic or eukaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the coding sequence.

A “**construct**”, e.g., a “nucleic acid construct” or “DNA construct” refers to a nucleic acid or nucleic acid sequence.

“**Derived from**” indicates a peptide or nucleotide sequence selected from within a given sequence. A peptide or nucleotide sequence derived from a named sequence may contain a small number of modifications relative to the parent sequence, in most cases representing deletion, replacement or insertion of less than about 15%, preferably less than about 10%, and in many cases less than about 5%, of amino acid residues or base pairs present in the parent sequence. In the case of DNAs, one DNA molecule is also considered to be derived from another if the two are capable of selectively hybridizing to one another. Typically, a derived peptide sequence will differ from a parent sequence by the replacement of up to 5 amino acids, in many cases up to 3 amino acids, and very often by 0 or 1 amino acids. Correspondingly, a derived nucleic acid sequence will differ from a parent sequence by the replacement of up to 15 bases, in many cases up to 9 bases, and very often by 0 - 3 bases. In some cases the amino acid(s) or base(s) is/are deleted rather than replaced.



**"Divalent"**, as that term is applied to ligands in this document, denotes a ligand which is capable of complexing with at least two protein molecules which contain ligand binding domains, to form a three (or greater number)-component complex.

**"Domain"** refers to a portion of a protein or polypeptide. In the art, domain may refer to a discrete 2° structure. However, as will be apparent from the context used herein, the term "domain" is not intended to be limited to a discrete folding domain. Rather, consideration of a polypeptide sequence as a "domain" in, e.g., a fusion protein herein, can be made simply by the observation that the polypeptide has a specific activity, function or source. Most domains described herein can be derived from proteins ranging from naturally occurring proteins to completely artificial sequences.

A **"DNA binding domain"** refers to a polypeptide which interacts, or binds, with a higher affinity to a nucleic acid having a particular nucleotide sequence relative to a nucleic acid having a different nucleotide sequence.

**"DNA recognition sequence"** means a DNA sequence which is capable of binding to one or more DNA-binding domains, e.g., of a transcription factor or an engineered polypeptide.

**"Endogenous"** refers to molecules which are naturally occurring in a cell, i.e. prior to the genetic engineering or infection of the cell.

**"Exogenous"** refers to molecules which are not naturally present in the cell, and which have been, e.g., introduced by transfection or transduction of the cell (or the parent cell thereof).

**"Gene"** refers to a nucleic acid molecule or sequence comprising an open reading frame and including at least one exon and (optionally) an intron sequence. The term "intron" refers to a DNA sequence present in a given gene which is not translated into protein and is generally found between exons.

**"Genetically engineered cells"** denotes cells which have been modified by the introduction of recombinant or heterologous nucleic acids (e.g. one or more DNA constructs or their RNA counterparts) and further includes the progeny of such cells which retain part or all of such genetic modification.

**"Heterologous"** as it relates to nucleic acid sequences such as coding sequences and control sequences, denotes sequences that are not normally joined together, and/or are not normally associated with a particular cell. Thus, a "heterologous" region of a nucleic acid construct is a segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with the other molecule in nature. For example, a heterologous region of a construct could include a coding sequence flanked by sequences not found in association with the coding sequence in nature. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g.,

synthetic sequences having codons different from the native gene). Similarly, in the case of a cell transduced with a nucleic acid construct which is not normally present in the cell, the cell and the construct would be considered mutually heterologous for purposes of this invention. Allelic variation or naturally occurring mutational events do not give rise to heterologous DNA, as used herein.

**"Interact"** as used herein is meant to include detectable interactions between molecules, such as can be detected using, for example, a yeast two hybrid assay or by immunoprecipitation. The term interact is also meant to include "binding" interactions between molecules. Interactions may be, for example, protein-protein, protein-nucleic acid, protein-small molecule or small molecule-nucleic acid in nature.

**"Ligand"** refers to any molecule which is capable of interacting with a corresponding protein or protein domain. A ligand can be naturally occurring, or the ligand can be partially or wholly synthetic. The term "modified ligand" refers to a ligand which has been modified such that it does not significantly interact with the naturally occurring receptor of the ligand in its non modified form. Ligands may be formulated and administered to cells or human or non-human animals as disclosed in the various patent documents cited herein.

A **"ligand binding domain"** is a domain which binds to a ligand or analogs or mimics thereof with measurable preference to binding to other materials. For the purpose of this document, DNA is not a ligand, and a DNA binding domain is not a ligand binding domain.

**"Minimal promoter"** refers to the minimal expression control sequence that is necessary for initiating transcription of a selected DNA sequence to which it is operably linked. For example, the term "minimal promoter" may be used to refer to a DNA sequence which is derived from a regulatory region upstream of a gene, contains a TATA box flanked upstream by about 20-30 base pairs and on its 3' end by ~100-300 bp, and which has little or no basal promoter activity, i.e., less than about 1% of the promoter activity observed with the full length regulatory region as determined by any measure of transcriptional activity.

The terms **"promoter"** and **"transcription control sequence"** further encompass "tissue specific" promoters and expression control sequences, i.e., promoters and expression control sequences which effect expression of the selected DNA sequence preferentially in specific cells (e.g., cells of a specific tissue). Gene expression occurs preferentially in a specific cell if expression in this cell type is significantly higher than expression in other cell types. The terms "promoter" and "expression control sequence" also encompass so-called "leaky" promoters and "expression control sequences", which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well. These terms also encompass non-tissue specific promoters and expression control sequences which are active in most cell types. Furthermore, a promoter or expression control sequence can be constitutive i.e. one which is active basally or inducible,

i.e., a promoter or expression control sequence which is active primarily in response to a stimulus. A stimulus can be, e.g., a molecule, such as a hormone, a cytokine, a heavy metal, phorbol esters, cyclic AMP (cAMP), or retinoic acid.

"**Nucleic acid**" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include, as equivalents, derivatives, variants and analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single (sense or antisense) and double-stranded polynucleotides.

"**Oligomerization**" and "**multimerization**", used interchangeably herein, refer to the association of two or more proteins which can be constitutive or inducible. Inducible oligomerization is mediated, in the practice of this invention, by the binding of each such protein to a common ligand. "Dimerization" refers to the association of two proteins. The formation of a tripartite (or greater) complex comprising proteins containing one or more FKBP domains together with one or more molecules of an FKBP ligand which is at least divalent (e.g. FK1012 or AP1510) is an example of such association or clustering. In cases where at least one of the proteins contains more than one ligand binding domain, e.g., whereat least one of the proteins contains three FKBP domains, the presence of a divalent ligand leads to the clustering of more than two protein molecules. Embodiments in which the ligand is more than divalent (e.g. trivalent) in its ability to bind to proteins bearing ligand binding domains also can result in clustering of more than two protein molecules. The formation of a tripartite complex comprising a protein containing at least one FRB domain, a protein containing at least one FKBP domain and a molecule of rapamycin is another example of such protein clustering. In certain embodiments of this invention, fusion proteins contain multiple FRB and/or FKBP domains. Complexes of such proteins may contain more than one molecule of rapamycin or a derivative thereof and more than one copy of one or more of the constituent proteins. Again, such multimeric complexes are still referred to herein as tripartite complexes to indicate the presence of the three types of constituent molecules, even if one or more are represented by multiple copies. The formation of complexes containing at least one divalent ligand and at least two molecules of a protein which contains at least one ligand binding domain may be referred to as "oligomerization" or "multimerization", or simply as "dimerization", "clustering" or association".

"**Operably linked**" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a transcription control sequence operably linked to a coding sequence permits expression of the coding sequence. The control sequence need not be contiguous with the coding sequence so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding

sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"**Protein**", "**polypeptide**" and "**peptide**" are used interchangeably herein when referring to a gene product, e.g., as may be encoded by a coding sequence.

5 A "**recombinant virus**" is a virus particle in which the packaged nucleic acid contains a heterologous portion.

A "**target gene**" is a nucleic acid of interest, the expression of which is modulated according to the methods of the invention. The target gene can be endogenous or exogenous and can integrate into a cell's genome, or remain episomal. The target gene can  
10 encode a protein or be a non coding nucleic acid, e.g, a nucleic acid which is transcribed into an antisense RNA or a ribozyme.

"**Transcription factor**" refers to any protein or modified form thereof that is involved in the initiation of transcription but which is not itself a part of the polymerase. Transcription factors are proteins or modified forms thereof, which interact preferentially with specific nucleic  
15 acid sequences, i.e., regulatory elements. Some transcription factors are active when they are in the form of a monomer. Alternatively, other transcription factors are active in the form of oligomers consisting of two or more identical proteins or different proteins (heterodimer). The factors have different actions during the transcription initiation: they may interact with other factors, with the RNA polymerase, with the entire complex, with activators, or with DNA.  
20 Transcription factors usually contain one or more transcription regulatory domains.

"**Transcription activation motifs**" as that phrase is used herein means a peptide motif of usually at least 6 amino acid residues which is either a transcription potentiating motif (i.e., it need not have a naturally occurring peptide sequence) or it is associated with a transcription activation domain, including, as non-limiting examples, the well-known "acidic",  
25 "glutamine-rich" and "proline-rich" motifs such as the K13 motif from p65, the OCT2 Q domain and the OCT2 P domain, respectively.

"**Transfection**" means the introduction of a naked nucleic acid molecule into a recipient cell. "**Infection**" refers to the process wherein a virus enters the cell in a manner whereby the genetic material of the virus can be expressed in the cell. A "productive  
30 infection" refers to the process wherein a virus enters the cell, is replicated, and then released from the cell (sometimes referred to as a "lytic" infection). "**Transduction**" encompasses the introduction of nucleic acid into cells by any means.

"**Transgene**" refers to a nucleic acid sequence which has been introduced into a cell. Daughter cells deriving from a cell in which a transgene has been introduced are also said to  
35 contain the transgene (unless it has been deleted). A transgene can encode, e.g., a polypeptide, partly or entirely heterologous to the animal or cell into which it is introduced, or comprises or is derived from an endogenous gene of the animal or cell into which it is

introduced, but which is designed to be inserted, or is inserted, into the recipient's genome in such a way as to alter that genome. (e.g., it is inserted at a location which differs from that of the natural gene). Alternatively, a transgene can also be present in an episome. A transgene can include one or more expression control sequences and any other nucleic acid, (e.g. intron), that may be necessary or desirable for optimal expression of a selected coding sequence.

**"Transient transfection"** refers to cases where exogenous DNA does not integrate into the genome of a transfected cell, e.g., where episomal DNA is transcribed into mRNA and translated into protein. A cell has been "stably transfected" with a nucleic acid construct when the nucleic acid construct has been integrated into the genome of that cell.

**"Wild-type"** means naturally occurring in a normal cell.

### Components of the system and additional details

The system, as employed in cells, comprises: (1) a recombinant nucleic acid construct encoding and capable of directing the expression of a chimeric transcription factor protein as described above; and, (2) a target gene construct containing a target gene and a transcription control sequence permitting transcription of the target gene under the direction of the chimeric transcription factor and in a ligand-dependent manner. The transcription control sequence comprises a DNA promoter sequence and one or more copies of a DNA recognition sequence to which the transcription factor is capable of binding ("recognizing"). In embodiments in which the chimeric transcription factor does not include a DNA binding domain, a second recombinant nucleic acid is included which encodes an accessory chimeric protein comprising at least one DNA binding domain and a ligand binding domain permitting ligand-dependent crosslinking with the chimeric transcription factor protein to form a two-hybrid-type chimeric transcription factor complex, capable of recognizing the target gene's transcription control sequence and stimulating transcription of the target gene in a ligand-dependent manner.

#### 1. Transcription Activation Domains

Chimeric transcription factors of this invention contain one or more copies of one or more p65 domains, one or more ligand binding domains, and optionally one or more copies of one or more transcription activation or potentiating domains. Such additional activation domains may be selected from peptide sequences of naturally occurring transcription factors such as the widely used transcription activation domain of Herpes Simplex Virus VP16 or human heat shock factor, may be derived from such sequences or may comprise a composite

transcription activation region. A composite transcription activation region consists of a continuous polypeptide region containing two or more reiterated or mutually heterologous component polypeptide portions. The component polypeptide portions comprise polypeptide sequences derived from at least two different proteins, polypeptide sequences  
5 from at least two non-adjacent portions of the same protein, polypeptide sequences which are not found so linked in nature (including reiterated copies of a polypeptide sequence) or non-naturally occurring peptide sequence. Preferably the activation domain or component peptide sequences thereof are selected or derived from peptide sequences endogenous to the cells or organism to be engineered.

10 NF-kb p65, preferably human, turned out to be an important source of transcription activation domains and motifs. p65(450-550) is a known transcription activation domain although methods and materials for using it as described herein have not been previously reported. We have found that extending the p65 peptide sequence to include sequence spanning p65 residues 361-450 leads to an unexpected increase in transcription activation.  
15 Moreover, a peptide sequence comprising all or a portion of p65(361-550), or peptide sequence derived therefrom, in combination with heterologous activation motifs, can yield surprising additional increases in the level of transcription activation. Incorporation of additional residues from p65, including those from 321-360 or 281-360, allows an even greater increase in transcription levels in certain systems. p65-based activation domains  
20 function across a broad range of promoters and have yielded increases in transcription levels six-fold, eight-fold and even 14-15-fold higher than obtained with tandem copies of VP16 which itself is widely recognized as a very potent activation domain.

While the resultant increases in activation potency are dramatic, p65-based transcription factors possess additional and unexpected characteristics. For instance, our  
25 p65-based activators appear to be less toxic to the engineered cells in comparison to VP16-based activators. This is clearly of profound practical significance in many applications. It is expected that recombinant DNA molecules encoding chimeric proteins which contain a peptide sequence comprising all or a portion of p65(361-550), p65(321-550) or p65(281-550), especially containing one or more portions of the sequence spanning residues 361 and  
30 450, or peptide sequence derived therefrom, will provide significant advantages for heterologous gene expression in its various contexts.

One class of p65-based chimeric transcription factors contain more than one copy of a p65-derived domain. Such proteins will typically contain two to six copies of a peptide sequence comprising all or a portion of p65(361-550), or peptide sequence derived  
35 therefrom. Such transcription factors may contain one or more ligand-binding domains to provide for regulation as described elsewhere herein, and in some cases additional other domains.

Chimeric transcription factors of this invention may contain, in addition to one or more copies of a p65 activation domain such as described above, one or more copies of one or more heterologous peptide sequences which potentiate the transcription activation potency of the transcription factor, as measured by any means. Inclusion of such motifs, including the so-called "glutamine-rich", "proline-rich" and "acidic" transcription activation motifs, in combination with a primary activation domain can result in extremely high levels of transcription.

A wide variety of transcription activation domains and motifs can be used in the practice of the present invention in conjunction with p65-based domains. Polypeptides which can function to activate transcription in eukaryotic cells are well known in the art. In particular, transcription activation domains have been described for many transcription factors and have been shown to retain their activation function when the transcription activation domain, or a suitable fragment thereof, is present within a fusion protein.

Activation domains can comprise naturally occurring or non-naturally occurring peptide sequences, so long as, either alone or in combination with other activation domains, they are capable of activating transcription. Any particular activation domain is preferably at least 6 amino acids in length. Naturally occurring activation domain subunits or motifs include portions of transcription factors, such as a thirty amino acid fragment of the C-terminus of VP16 (amino acids 461-490), referred to herein as "Vc". Other activation domain subunits are derivatives of naturally occurring peptides. For example, the replacement of one amino acid of a naturally occurring activation unit by another may further increase activation. An example of such an activation unit is a derivative of an eight amino acid peptide of VP16, the derivative having the amino acid sequence DFDLMLG.

Yet other activation units are entirely synthetic. It is known, for example, that certain random alignments of acidic amino acids are capable of activating transcription.

It is well known in the art that certain transcription factors are active only in specific cell types. By using tissue specific activation domains, it is possible to design a transcription factor having a certain tissue specificity.

One source of polypeptide motifs for use in conjunction with p65-based activation domains is the herpes simplex virus virion protein 16 (referred to herein as VP16, the amino acid sequence of which is disclosed in Triezenberg, S.J. et al. (1988) Genes Dev. 2:718-729). In one embodiment, an activation domain corresponding to about 127 of the C-terminal amino acids of VP16 is used. For example, a polypeptide having amino acid residues 208-335 can be used as an auxiliary activation domain. In another embodiment, at least one copy of about 11 amino acids from the C-terminal region of VP16 which retain transcription activation ability is used as an additional activation domain. Preferably, an oligomer of this region (i.e., about 22 amino acids) is used. Suitable C-terminal peptide portions of VP16 are

described in Seipel, K. et al. (EMBO J. (1992) 13:4961-4968). VP16-derived transcription activation domains have been used successfully in many of the different regulated expression systems referred to herein.

Another example of an acidic activation domain is provided in residues 753-881 of GAL4. A preferred activation domain to be used in conjunction with a p65 domain is a domain from a human heat shock factor transcriptional activator (HSF domain), in particular, residues 406-530. Heat shock factor activation domains are more fully described in USSN 09/262,600, filed 3/4/99, the full contents of which are incorporated herein by reference.

Other illustrative activation domains and motifs of human origin include the activation domain of human CTF, the 18 amino acid (NFLQLPQQTQGALLTSQP) glutamine rich region of Oct-2, the N-terminal 72 amino acids of p53, the SYGQQS repeat in Ewing sarcoma gene and an 11 amino acid (535-545) acidic rich region of Rel A protein. Various additional activation domains, motifs and chimeric transcription factors are provided in the examples which follow. See also USSN 08/920,610 (ARIAD 363-A), the contents of which are incorporated herein by reference, especially for additional information concerning sources of activation domains and motifs that may be used in combination with p65 domains in the chimeric transcription domains of this invention.

## **2. Ligand binding domains**

The chimeric transcription factors contain at least one p65 domain and one ligand binding domain, but function, in the various embodiments, through different molecular mechanisms.

### **A. Dimerization-based systems**

In certain embodiments, the ligand binding domain permits ligand-mediated cross-linking of the chimeric transcription factor with a second fusion protein (which contains at least one ligand binding domain and DNA binding domain). In these cases, the ligand is at least divalent and functions as a dimerizing agent by binding to the two fusion proteins and forming a cross-linked heterodimeric complex which activates target gene expression. See e.g. WO 94/18317, WO 96/20951, WO 96/06097, WO 97/31898, WO 96/41865, and PCT US98/17723, the contents of which are incorporated herein by reference.

In other embodiments, the ligand binding event is thought to result in an allosteric change in the chimeric transcription factor leading to binding of the fusion protein to a target DNA sequence [see e.g. US 5,654,168 and 5,650,298 (tet systems), and WO 93/23431 and WO 98/18925 (RU486-based systems)] or to another protein [see e.g. WO 96/37609 and



WO 97/38117 (ecdysone/RXR-based systems)], in either case, modulating target gene expression.

In the cross-linking-based dimerization systems the fusion proteins can contain one or more ligand binding domains (in some cases containing two, three or four such domains) and  
5 can further contain one or more additional domains, heterologous thereto, including e.g. a DNA binding domain, transcription activation domain, etc.

In general, any ligand/ligand binding domain pair may be used in such systems. For example, ligand binding domains may be derived from an immunophilin such as an FKBP, cyclophilin, FRB domain, hormone receptor protein, antibody, etc., so long as a ligand is  
10 known or can be identified for the ligand binding domain.

For the most part, the receptor domains will be at least about 50 amino acids, and fewer than about 350 amino acids, usually fewer than 200 amino acids, either as the natural domain or truncated active portion thereof. Preferably the binding domain will be small (<25 kDa, to allow efficient transfection in viral vectors), monomeric, nonimmunogenic, and should  
15 have synthetically accessible, cell permeant, nontoxic ligands as described above.

Preferably the ligand binding domain is for (i.e., binds to) a ligand which is not itself a gene product (i.e., is not a protein), has a molecular weight of less than about 5 kD and preferably less than about 3 kD, and is cell permeant. In many cases it will be preferred that the ligand does not have an intrinsic pharmacologic activity or toxicity which interferes with its  
20 use as a transcription regulator.

The DNA sequence encoding the ligand binding domain can be subjected to mutagenesis for a variety of reasons. The mutagenized ligand binding domain can provide for higher binding affinity, allow for discrimination by a ligand between the mutant and naturally occurring forms of the ligand binding domain, provide opportunities to design ligand-  
25 ligand binding domain pairs, or the like. The change in the ligand binding domain can involve directed changes in amino acids known to be involved in ligand binding or with ligand-dependent conformational changes. Alternatively, one may employ random mutagenesis using combinatorial techniques. In either event, the mutant ligand binding domain can be expressed in an appropriate prokaryotic or eukaryotic host and then screened for desired  
30 ligand binding or conformational properties. Examples involving FKBP, cyclophilin and FRB domains are disclosed in detail in WO 94/18317, WO 96/06097, WO 97/31898 and WO 96/41865). Illustrative of this situation is to modify FKBP12's Phe36 to Ala and/or Asp37 to Gly or Ala to accommodate a substituent at positions 9 or 10 of FK506 or FK520. In particular, mutant FKBP12 moieties which contain Val, Ala, Gly, Met or other small amino acids in place  
35 of one or more of Tyr26, Phe36, Asp37, Tyr82 and Phe99 are of particular interest as receptor domains for FK506-type and FK-520-type ligands containing modifications at C9

and/or C10. Illustrative mutations of current interest in FKBP domains also include the following:

F36A	Y26V	F46A	W59A
F36V	Y26S	F48H	H87W
F36M	D37A	F48L	H87R
F36S	I90A	F48A	F36V/F99A
F99A	I91A	E54A/F36V/F99G	F99G
F46H	E54K/F36M/F99A	Y26A	F46L
V55A	F36M/F99G		

5 Table 1: Entries identify the native amino acid by single letter code and sequence position, followed by the replacement amino acid in the mutant. Thus, F36V designates a human FKBP12 sequence in which phenylalanine at position 36 is replaced by valine. F36V/F99A indicates a double mutation in which phenylalanine at positions 36 and 99 are replaced by valine and alanine, respectively.

10 Illustrative examples of rapamycin-binding domains are those which include an approximately 89-amino acid rapamycin-binding domain from FRAP, e.g., containing residues 2025-2113 of human FRAP. Another preferred portion of FRAP is a 93 amino acid fragment consisting of amino acids 2021-2113. Similar considerations apply to the generation of

15 mutant FRAP-derived domains which bind preferentially to rapamycin analogs (rapalogs) containing modifications (i.e., are 'bumped') relative to rapamycin in the FRAP-binding effector domain. For example, one may obtain preferential binding using rapalogs bearing substituents other than -OMe at the C7 position with FRBs based on the human FRAP FRB peptide sequence but bearing amino acid substitutions for one or more of the residues

20 Tyr2038, Phe2039, Thr2098, Gln2099, Trp2101 and Asp2102. Exemplary mutations include Y2038H, Y2038L, Y2038V, Y2038A, F2039H, F2039L, F2039A, F2039V, D2102A, T2098A, T2098N, T2098L, and T2098S. Rapalogs bearing substituents other than -OH at C28 and/or substituents other than =O at C30 may be used to obtain preferential binding to FRAP proteins bearing an amino acid substitution for Glu2032. Exemplary mutations include

25 E2032A and E2032S. Proteins comprising an FRB containing one or more amino acid replacements at the foregoing positions, libraries of proteins or peptides randomized at those positions (i.e., containing various substituted amino acids at those residues), libraries randomizing the entire protein domain, or combinations of these sets of mutants are made using the procedures described above to identify mutant FRAPs that bind preferentially to

30 bumped rapalogs. See, for example, USSN 09/012,097, the contents of which are incorporated herein by reference.

Other macrolide binding domains useful in the present invention, including mutants thereof, are described in the art. See, for example, WO96/41865, WO96/13613, WO96/06111, WO96/06110, WO96/06097, WO96/12796, WO95/05389, WO95/02684, WO94/18317, each of which is expressly incorporated by reference herein.

5       The ability to employ *in vitro* mutagenesis or combinatorial modifications of sequences encoding proteins allows for the production of libraries of proteins which can be screened for binding affinity for different ligands. For example, one can totally randomize a sequence of 1 to 5, 10 or more codons, at one or more sites in a DNA sequence encoding a binding protein, make an expression construct and introduce the expression construct into a unicellular  
10   microorganism, and develop a library. One can then screen the library for binding affinity to one or desirably a plurality of ligands. The best affinity sequences which are compatible with the cells into which they would be introduced can then be used as the ligand binding domain. The ligand would be screened with the host cells to be used to determine the level of binding of the ligand to endogenous proteins. A binding profile could be defined weighting  
15   the ratio of binding affinity to the mutagenized binding domain with the binding affinity to endogenous proteins. Those ligands which have the best binding profile could then be used as the ligand. Phage display techniques, as a non-limiting example, can be used in carrying out the foregoing.

      In other embodiments, antibody subunits, e.g. heavy or light chain, particularly  
20   fragments, more particularly all or part of the variable region, or fusions of heavy and light chain to create single chain antibodies, can be used as the ligand binding domain. Antibodies can be prepared against haptenic molecules which are physiologically acceptable and the individual antibody subunits screened for binding affinity. The cDNA encoding the subunits can be isolated and modified by deletion of the constant region, portions of the variable  
25   region, mutagenesis of the variable region, or the like, to obtain a binding protein domain that has the appropriate affinity for the ligand. In this way, almost any physiologically acceptable haptenic compound can be employed as the ligand or to provide an epitope for the ligand. Instead of antibody units, natural receptors can be employed, where the binding domain is known and there is a useful ligand for binding. In yet another embodiment of the  
30   invention, the DNA binding unit is linked to more than one ligand binding domain. For example, a DNA binding domain can be linked to at least 2, 3, 4, or 5 ligand binding domains. A DNA binding domain can also be linked to at least 5 ligand binding domains or any number of ligand binding domains. In such embodiments, the ligand binding domains can be, by illustration, linked to each other in a linear array, by linking the NH<sub>2</sub>-terminus of one ligand  
35   binding domain to the COOH-terminus of another ligand binding domain. Thus, more than one molecule of a chimeric transcription factor can be cross-linked to a single DNA binding domain in the presence of a divalent ligand.

## B. Allosteric-based systems

As mentioned previously, ligand-dependent transcription regulation switches based on allosteric changes in a chimeric transcription factor are also useful in practicing the subject invention. One such switch employs a deletion mutant of the human progesterone receptor which no longer binds progesterone or any known endogenous steroid but can be activated by the orally active progesterone antagonist RU486, described, e.g. in Wang et al. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:8180. The transcription factor in this system generally consists of a ligand binding domain for binding RU486, a DNA binding domain such as GAL4 and an activation domain, typically VP16. Activation was demonstrated, e.g. in cells transplanted into mice using doses of RU486 (5-50 µg/kg) considerably below the usual dose for inducing abortion in humans (10 mg/kg). However, according to the art describing this system, the induction ratio in culture and in animals was rather low.

Another such system is referred to as the ecdysone inducible system. Early work demonstrated that fusing the *Drosophila* steroid ecdysone (Ec) receptor (EcR) Ec-binding domain to heterologous DNA binding and activation domains, such as *E. coli* *lexA* and herpesvirus VP16 permits ecdysone-dependent activation of target genes downstream of appropriate binding sites (Christopherson et al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:6314). An improved ecdysone regulation system has been developed, using the DNA binding domain of the EcR itself. In this system, the regulating transcription factor is provided as two proteins: (1) a truncated, mutant EcR fused to herpes VP16 and (2) the mammalian homolog (RXR) of Ultraspiracle protein (USP), which heterodimerizes with the EcR (No et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:3346). In this system, because the DNA binding domain was also recognized by a human receptor (the human farnesoid X receptor), it was altered to a site recognized only by the mutant EcR. Thus, the invention provides an ecdysone inducible system, in which a truncated mutant EcR is fused to at least one subunit of a transcription activator of the invention. The transcription factor further comprises USP, thereby providing high level induction of transcription of a target gene having the EcR target sequence, dependent on the presence of ecdysone.

In another embodiment, the inducible system comprises the *E. coli* tet repressor (TetR), which binds to tet operator (tetO) sequences upstream of target genes. In the presence of tetracycline, or an analog, which bind to tetR, DNA binding is abolished and thus transactivation is abolished. This system, in which the TetR had previously been linked to transcription activation domains, e.g. from VP16, is generally referred to as an allosteric "off-switch" described by Gossen and Bujard (Proc. Natl. Acad. Sci. U.S.A. (1992) 89:5547) and in U.S. Patents 5,464,758; 5,650,298; and 5,589,362 by Bujard et al. Furthermore, depending on the concentration of the antibiotic in the culture medium (0-1 µg/ml), target

gene expression can be regulated over concentrations up to several orders of magnitude. Thus, the system not only allows differential control of the activity of an individual gene in eukaryotic cells but also is suitable for creation of "on/off" situations for such genes in a reversible way. This system provides target gene expression in the absence of tetracycline or an analog. Thus, the invention described herein provides a method for obtaining even stronger transcription induction of a target gene, which is regulatable by the tetracycline system or other inducible DNA binding domain.

In another embodiment, a "reverse" Tet system is used, again based on a DNA binding domain that is a mutant of the E. coli TetR, but which binds to TetO in the presence of Tet. Additional information on mutated tetR-based systems is provided above and in patent documents cited previously. Thus, the invention described herein provides a method for obtaining even stronger transcription induction of a target gene in the presence of tetracycline or an analog thereof from a very low background in the absence of tetracycline.

### 3. Ligands of the invention

In various embodiments where the a ligand binding domain for the ligand is endogenous to the cells to be engineered, it is desirable to use a ligand which preferentially binds to a modified ligand binding domain relative to a naturally occurring peptide sequence, e.g., from this the modified domain was derived. This approach can avoid untoward intrinsic activities of the ligand. Significant guidance and illustrative examples toward that end are provided in the various references cited herein.

#### A. Cross-linking/dimerization systems

Any ligand for which a binding protein or ligand binding domain is known or can be identified may be used in combination with such ligand binding domain in carrying out this invention.

Extensive guidance and examples are provided in WO 94/18317 for ligands and other components useful for cross-linked oligomerization-based systems. Systems based on ligands for an immunophilin such as FKBP, a cyclophilin, and/or FRB domain are of special interest. Illustrative examples of ligand binding domain/ligand pairs that may be used for cross-linking include, but are not limited to: FKBP/FK1012, FKBP/synthetic divalent FKBP ligands (see WO 96/06097 and WO 97/31898), FRB/rapamycin or analogs thereof: FKBP (see e.g., WO 93/33052, WO 96/41865 and Rivera et al, "A humanized system for pharmacologic control of gene expression", Nature Medicine 2(9):1028-1032 (1997)), cyclophilin/cyclosporin (see e.g. WO 94/18317), FKBP/FKCsA/cyclophilin (see e.g. Belshaw et al, 1996, PNAS 93:4604-4607), DHFR/methotrexate (see e.g. Licitra et al, 1996, Proc. Natl. Acad. Sci. USA 93:12817-12821), and DNA gyrase/coumermycin (see e.g. Farrar

et al, 1996, Nature 383:178-181). Numerous variations and modifications to ligands and ligand binding domains, as well as methodologies for designing, selecting and/or characterizing them, which may be adapted to the present invention are disclosed in the cited references.

5    B.     *Allostery-based systems*

For additional guidance on ligands for other systems which may be adapted to this invention, see e.g. (Gossen and Bujard Proc. Natl. Acad. Sci. U.S.A. 1992 89:5547, and US Patent Nos. 5654168, 5650298, 5589362 and 5464758 (TetR/tetracycline), Wang et al, 1994, Proc. Natl. Acad. Sci. USA 91:8180-8184 (progesterone receptor/RU486), and No et al, 1996, Proc. Natl. Acad. Sci. USA 93:3346-3351 (ecdysone receptor/ecdysone).

10

#### 4. DNA-binding domains

Regulated expression systems relevant to this invention involve the use of a protein containing a DNA binding domain to selectively target a desired gene for expression. In systems based on ligand-mediated cross-linking, the DNA binding domain can be provided in a fusion protein with one or more ligand binding domains. In certain embodiments based on allosteric mechanisms, e.g. the tetR and progesterone-R-based systems, the transcription activation domain is provided as part of a fusion protein which further comprises a DNA-binding domain.

15

Various DNA binding domains may be incorporated into the design of the chimeric transcription factor (or companion fusion protein) so long as a corresponding DNA "recognition" sequence is known or can be identified to which the domain is capable of binding. One or more copies of the recognition sequence are incorporated into the transcription control sequence of the target gene construct. Peptide sequence of human origin is often preferred, where available, for uses in human gene therapy. Composite DNA binding domains provide one means for achieving novel sequence specificity for the protein-DNA binding interaction. An illustrative composite DNA binding domain containing component peptide sequences of human origin is ZFHD-1 which is described in detail below. Individual DNA-binding domains may be further modified by mutagenesis to decrease, increase, or change the recognition specificity of DNA binding. These modifications could be achieved by rational design of substitutions in positions known to contribute to DNA recognition (often based on homology to related proteins for which explicit structural data are available). For example, in the case of a homeodomain, substitutions can be made in amino acids in the N-terminal arm, first loop, second helix, and third helix known to contact DNA. In zinc fingers, substitutions can be made at selected positions in the DNA recognition helix. Alternatively, random methods, such as selection from a phage display library could be used to identify altered domains with increased affinity or altered specificity. Individual DNA-binding domains

20

25

30

35

may be further modified by mutagenesis to decrease, increase, or change the recognition specificity of DNA binding. These modifications could be achieved by rational design of substitutions in positions known to contribute to DNA recognition (often based on homology to related proteins for which explicit structural data are available). For example, in the case of a homeodomain, substitutions can be made in amino acids in the N-terminal arm, first loop, second helix, and third helix known to contact DNA. In zinc fingers, substitutions can be made at selected positions in the DNA recognition helix. Alternatively, random methods, such as selection from a phage display library can be used to identify altered domains with increased affinity or altered specificity.

One type of DNA binding domain of interest is a DNA binding domain which binds to a characteristic DNA sequence in a ligand-dependent manner. This sort of DNA binding domain is exemplified by the Tet repressor (tetR) and mutated versions thereof which are discussed in detail elsewhere herein and in various cited references.

See the various patent and scientific documents cited herein, and in particular, WO 96/20951, WO 94/18317 and USSN 60/084819 for additional guidance on DNA binding domains and their use in such systems, as well as on other aspects of this invention.

## 5. Regulatory domains

Domains which regulate the transcriptional activity of p65-containing chimeric transcription factors may be included in the chimeric transcription factors of this invention. Such domains may be, for example, multimerization domains, as described in Natesan USSN 09/140,149. One example of such a multimerization domain is the trimerization domain which spans amino acids 126-217 of human HSF1. This domain is required for DNA binding to the wild-type heat shock transcription factor (Morimoto 1998, Genes and Development 12:3788-3796) and may be used e.g., as previously disclosed, to recruit additional activation domains to the promoter. Optionally, the regulatory domain of HSF1, comprising amino acids 201-371, may be used in the chimeric transcription factors of this invention. In some embodiments, a smaller fragment comprising amino acids 221-310 may be used. In yet other embodiments, a minimal regulatory domain comprising amino acids 300-310 which contains the serine phosphorylation sites at positions 303 and 307 may be used. If desired, one or more of the serines at positions 303 and 307 in the context of the full length or minimal regulatory domain may be replaced by different amino acids. As described in Green et al., supra, the regulatory domain negatively regulates the transcriptional activity of the HSF domain in the wild type transcription factor as well as in fusion proteins containing the GAL4 DNA binding domain. If lower levels of transcriptional activation are required or desired, this domain may be fused to the p65 domain alone or in combination with other potentiating or regulatory domains.

Another regulatory domain which may be used in combination with a p65 domain in the practice of this invention comprises the peptide sequence spanning amino acids 280-360 of human NF-kb p65. In constructs expressing p65(281-550), the regulatory region stabilizes the activation domain (361-550) and allows higher levels of transcription factor to be expressed in the cell.

## 6. Additional domains

Additional domains may be included in chimeric proteins of this invention. For example, the chimeric proteins may contain a nuclear localization sequence which provides for translocation of the protein to the nucleus. Typically a nuclear localization sequence has a plurality of basic amino acids, referred to as a bipartite basic repeat (reviewed in Garcia-Bustos *et al*, Biochimica et Biophysica Acta (1991) 1071, 83-101). This sequence can appear in any portion of the molecule internal or proximal to the N- or C-terminus and results in the chimeric protein being localized inside the nucleus.

The chimeric proteins may include domains that facilitate their purification, *e.g.* "histidine tags" or a glutathione-S-transferase domain. They may include "epitope tags" encoding peptides recognized by known monoclonal antibodies for the detection of proteins within cells or the capture of proteins by antibodies *in vitro*.

Transcription factors can be tested for activity *in vivo* using a simple assay (F.M. Ausubel *et al.*, Eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (John Wiley & Sons, New York, 1994); de Wet *et al.*, *Mol. Cell Biol.* 7:725 (1987)). The *in vivo* assay requires a plasmid containing and capable of directing the expression of a recombinant DNA sequence encoding the transcription factor. The assay also requires a plasmid containing a reporter gene, *e.g.*, the luciferase gene, the chloramphenicol acetyl transferase (CAT) gene, secreted alkaline phosphatase or the human growth hormone (hGH) gene, linked to a binding site for the transcription factor. The two plasmids are introduced into host cells which normally do not produce interfering levels of the reporter gene product. A second group of cells, which also lack both the gene encoding the transcription factor and the reporter gene, serves as the control group and receives a plasmid containing the gene encoding the transcription factor and a plasmid containing the test gene without the binding site for the transcription factor.

The production of mRNA or protein encoded by the reporter gene is measured. An increase in reporter gene expression not seen in the controls indicates that the transcription factor is a positive regulator of transcription. If reporter gene expression is less than that of the control, the transcription factor is a negative regulator of transcription.

Optionally, the assay may include a transfection efficiency control plasmid. This plasmid expresses a gene product independent of the test gene, and the amount of this gene product indicates roughly how many cells are taking up the plasmids and how



efficiently the DNA is being introduced into the cells. Additional guidance on evaluating chimeric proteins of this invention is provided below.

## 7. Promoters

5        The activation domains described herein may be used in combination with any promoter that will direct their expression in mammalian cells. The promoter may be a strong promoter, such as the human CMV promoter, or a weaker promoter, such as a promoter for an endogenous human gene. Other promoters which may be used include, but are not limited to, the Rous Sarcoma Virus (RSV) promoter, the retroviral LTR from Murine Moloney  
10    Leukemia Virus (MMLV), the muscle creatine kinase (MCK) enhancer, the SV40 promoter, and the CMV enhancer from the major immediate early gene. Genbank accession numbers for the above promoters are given in the table below.

Promoter	Genbank Accession Number
CMV	AF067197
RSV	M83237
MMLV LTR	M77239
SV40	U47120
CMV enhancer for MIE gene	K03104
MCK enhancer	X67536

15        In many cases, the selection of promoter will depend upon the configuration of the transcription factor, i.e., which activation domain is used and whether any potentiating or stabilizing domains are added. Thus, a relatively weaker promoter such as RSV or a promoter for an endogenous human gene may be selected for use with a stabilizing domain, for example, amino acids 280-320 of p65 or the regulatory domain of HSF. In this type of  
20    configuration, the lower level of transcription which is normally driven by a weaker promoter may be compensated for by a more highly expressed, stable transcription factor.

Exemplary configurations of chimeric transcription factors of this invention include an RSV promoter, a p65 (320-550) activation domain and a regulatory domain from HSF. Other examples include (i) a CMV promoter and a p65(361-550) activation domain, (ii) a CMV  
25    promoter, a p65(361-550) activation domain and a HSF(431-529) activation domain, (iii) an MCK enhancer, a p65 (380-320) regulatory domain and a p65(361-550) activation domain and (iv) an RSV promoter, a p65(281-551) activation domain and a HSF(406-530) activation

domain. Any other combination of promoters, activation domains and potentiating domains contemplated by the practitioner may be used in the practice of this invention.

## 8. Transcription factors, additional comments.

5 In engineering cells for or in whole animals in accordance with this invention, it will often be preferred, and in some cases required, that the various domains or subdomains of the chimeric transcription factors be derived from proteins of the same species as the host cell. Thus, for genetic engineering of human cells, it is often preferred that component peptide sequences of human origin be used in some or all cases, rather than of bacterial, yeast or  
10 other non-human source. Transcription factor constructs generally contain (1) a promoter region consisting minimally of a TATA box and initiator sequence but optionally including other transcription factor binding sites; (2) DNA sequence encoding the desired transcription factor, including sequences that promote the initiation and termination of translation, if appropriate; (3) an optional sequence consisting of a splice donor, splice acceptor, and  
15 intervening intron DNA; and (4) a sequence directing cleavage and polyadenylation of the resulting RNA transcript. The practitioner may select a conventional promoter such as the widely used hCMV promoter region

It will be preferred in certain embodiments, especially where DNA is introduced into an animal for uptake by cells *in situ*, that the transcription factors be expressed in a cell-specific  
20 or tissue-specific manner. Such specificity of expression may be achieved by operably linking one or more of the DNA sequences encoding the chimeric protein(s) to a cell-type specific transcriptional regulatory sequence (e.g. promoter/enhancer). Numerous cell-type specific transcriptional regulatory sequences are known. Others may be obtained from genes which are expressed in a cell-specific manner. For example, constructs for expressing the  
25 chimeric proteins may contain regulatory sequences derived from genes known to exhibit expression in selected tissues. See e.g. PCT/US95/10591 (attorney dkt ARIAD 349), especially pp. 36-37.

## 9. Target gene constructs

30 A DNA construct that enables transcription of a target gene to be regulated by a transcription factor in accordance with this invention comprises a DNA molecule which includes a synthetic transcription unit typically consisting of: (1) one copy or multiple copies of a DNA sequence recognized with high-affinity by the transcription factor or one or more of its component DNA binding domains; (2) a promoter sequence consisting minimally of a TATA box and initiator  
35 sequence but optionally including other transcription factor binding sites; (3) sequence encoding the desired product, including sequences that promote the initiation and termination of translation, if appropriate; (4) an optional sequence consisting of a splice donor, splice

acceptor, and intervening intron DNA; and (5) a sequence directing cleavage and polyadenylation of the resulting RNA transcript. Typically the gene construct contains a copy of the target gene to be expressed, operably linked to a transcription control sequence comprising a minimal promoter and one or more copies of a DNA recognition sequence responsive to the transcription factor.

#### (a) Target genes

A wide variety of genes can be employed as the target gene, including genes that encode a therapeutic protein, antisense sequence or ribozyme of interest. The target gene can be any sequence of interest which provides a desired phenotype. It can encode a surface membrane protein, a secreted protein, a cytoplasmic protein, or there can be a plurality of target genes encoding different products. The target gene may be an antisense sequence which can modulate a particular pathway by inhibiting a transcriptional regulation protein or turn on a particular pathway by inhibiting the translation of an inhibitor of the pathway. The target gene can encode a ribozyme which may modulate a particular pathway by interfering, at the RNA level, with the expression of a relevant transcriptional regulator or with the expression of an inhibitor of a particular pathway. The proteins which are expressed, singly or in combination, can involve homing, cytotoxicity, proliferation, immune response, inflammatory response, clotting or dissolving of clots, hormonal regulation, etc. The proteins expressed may be naturally-occurring proteins, mutants of naturally-occurring proteins, unique sequences, or combinations thereof.

Various secreted products include hormones, such as insulin, human growth hormone, glucagon, pituitary releasing factor, ACTH, melanotropin, relaxin, *etc.*; growth factors, such as EGF, IGF-1, TGF- $\alpha$ , - $\beta$ , PDGF, G-CSF, M-CSF, GM-CSF, FGF, erythropoietin, thrombopoietin, megakaryocytic stimulating and growth factors, *etc.*; interleukins, such as IL-1 to -13; TNF- $\alpha$  and - $\beta$ , *etc.*; and enzymes and other factors, such as tissue plasminogen activator, members of the complement cascade, perforins, superoxide dismutase, coagulation factors, antithrombin-III, Factor VIIIc, vWF, Factor IX, alpha-anti-trypsin, protein C, protein S, endorphins, dynorphin, bone morphogenetic protein, CFTR, *etc.*

The gene can encode a naturally-occurring surface membrane protein or a protein made so by introduction of an appropriate signal peptide and transmembrane sequence. Various such proteins include homing receptors, *e.g.* L-selectin (Mel-14), blood-related proteins, particularly having a kringle structure, *e.g.* Factor VIIIc, Factor VIIIvW, hematopoietic cell markers, *e.g.* CD3, CD4, CD8, B cell receptor, TCR subunits  $\alpha$ ,  $\beta$ ,  $\gamma$  or  $\delta$ , CD10, CD19, CD28, CD33, CD38, CD41, *etc.*, receptors, such as the interleukin

receptors IL-2R, IL-4R, *etc.*, channel proteins, for influx or efflux of ions, *e.g.* H<sup>+</sup>, Ca<sup>2+</sup>, K<sup>+</sup>, Na<sup>+</sup>, Cl<sup>-</sup>, *etc.*, and the like; CFTR, tyrosine activation motif, zap-70, *etc.*

Proteins may be modified for transport to a vesicle for exocytosis. By adding the sequence from a protein which is directed to vesicles, where the sequence is modified proximal to one or the other terminus, or situated in an analogous position to the protein source, the modified protein will be directed to the Golgi apparatus for packaging in a vesicle. This process in conjunction with the presence of the chimeric proteins for exocytosis allows for rapid transfer of the proteins to the extracellular medium and a relatively high localized concentration.

Also, intracellular proteins can be of interest, such as proteins in metabolic pathways, regulatory proteins, steroid receptors, transcription factors, *etc.*, depending upon the nature of the host cell. Some of the proteins indicated above can also serve as intracellular proteins.

By way of further illustration, in T-cells, one may wish to introduce genes encoding one or both chains of a T-cell receptor. For B-cells, one could provide the heavy and light chains for an immunoglobulin for secretion. For cutaneous cells, *e.g.* keratinocytes, particularly stem cells keratinocytes, one could provide for protection against infection, by secreting alpha, beta or gamma interferon, antichemotactic factors, proteases specific for bacterial cell wall proteins, *etc.*

In addition to providing for expression of a gene having therapeutic value, there will be many situations where one may wish to direct a cell to a particular site. The site can include anatomical sites, such as lymph nodes, mucosal tissue, skin, synovium, lung or other internal organs or functional sites, such as clots, injured sites, sites of surgical manipulation, inflammation, infection, *etc.* By providing for expression of surface membrane proteins which will direct the host cell to the particular site by providing for binding at the host target site to a naturally-occurring epitope, localized concentrations of a secreted product can be achieved. Proteins of interest include homing receptors, *e.g.* L-selectin, GMP140, CLAM-1, *etc.*, or addressins, *e.g.* ELAM-1, PNA<sub>D</sub>, LNA<sub>D</sub>, *etc.*, clot binding proteins, or cell surface proteins that respond to localized gradients of chemotactic factors. There are numerous situations where one would wish to direct cells to a particular site, where release of a therapeutic product could be of great value.

#### **(b) Minimal Promoters.**

Minimal promoters may be selected from a wide variety of known sequences, including promoter regions from fos, hCMV, SV40 and IL-2, among many others. Illustrative examples are provided which use a minimal CMV promoter or a minimal IL2 gene promoter (-72 to +45 with respect to the start site; Siebenlist et al., MCB 6:3042-3049, 1986).

Promoter	Genbank Accession Number
CMV	AF067197
fos	M16287
IL-2	U57613
SV40	U47120
CMV enhancer for MIE gene	K03104

### (c) DNA recognition sequences.

Recognition sequences for a wide variety of DNA-binding domains are known. DNA  
 5 recognition sequences for other DNA binding domains may be determined experimentally. In  
 the case of a composite DNA binding domain, DNA recognition sequences can be determined  
 experimentally, as described below, or the proteins can be manipulated to direct their  
 specificity toward a desired sequence. A desirable nucleic acid recognition sequence for a  
 composite DNA binding domain consists of a nucleotide sequence spanning at least ten,  
 10 preferably eleven, and more preferably twelve or more bases. The component binding  
 portions (putative or demonstrated) within the nucleotide sequence need not be fully  
 contiguous; they may be interspersed with "spacer" base pairs that need not be directly  
 contacted by the chimeric protein but rather impose proper spacing between the nucleic acid  
 subsites recognized by each module. These sequences should not impart expression to  
 15 linked genes when introduced into cells in the absence of the engineered DNA-binding  
 protein.

To identify a nucleotide sequence that is recognized by a chimeric protein containing a  
 DNA-binding region, preferably recognized with high affinity (dissociation constant  $10^{-11}$  M  
 or lower are especially preferred), several methods can be used. If high-affinity binding sites  
 20 for individual subdomains of a composite DNA-binding region are already known, then these  
 sequences can be joined with various spacing and orientation and the optimum configuration  
 determined experimentally (see below for methods for determining affinities). Alternatively,  
 high-affinity binding sites for the protein or protein complex can be selected from a large pool  
 of random DNA sequences by adaptation of published methods (Pollock, R. and Treisman,  
 25 R., 1990, A sensitive method for the determination of protein-DNA binding specificities. *Nucl.*  
*Acids Res.* 18, 6197-6204). Bound sequences are cloned into a plasmid and their precise  
 sequence and affinity for the proteins are determined. From this collection of sequences,  
 individual sequences with desirable characteristics (*i.e.*, maximal affinity for composite

protein, minimal affinity for individual subdomains) are selected for use. Alternatively, the collection of sequences is used to derive a consensus sequence that carries the favored base pairs at each position. Such a consensus sequence is synthesized and tested to confirm that it has an appropriate level of affinity and specificity.

- 5 The target gene constructs may contain multiple copies of a DNA recognition sequence. For instance, the constructs may contain 5, 8, 10 or 12 recognition sequences for GAL4 or for ZFHD1.

**(d) Determination of binding affinity.**

- 10 A number of well-characterized assays are available for determining the binding affinity, usually expressed as dissociation constant, for DNA-binding proteins and the cognate DNA sequences to which they bind. These assays usually require the preparation of purified protein and binding site (usually a synthetic oligonucleotide) of known concentration and specific activity. Examples include electrophoretic mobility-shift assays, 15 DNaseI protection or "footprinting", and filter-binding. These assays can also be used to get rough estimates of association and dissociation rate constants. These values may be determined with greater precision using a BIAcore instrument. In this assay, the synthetic oligonucleotide is bound to the assay "chip," and purified DNA-binding protein is passed through the flow-cell. Binding of the protein to the DNA immobilized on the chip is measured 20 as an increase in refractive index. Once protein is bound at equilibrium, buffer without protein is passed over the chip, and the dissociation of the protein results in a return of the refractive index to baseline value. The rates of association and dissociation are calculated from these curves, and the affinity or dissociation constant is calculated from these rates. Binding rates and affinities for the high affinity composite site may be compared with the values obtained 25 for subsites recognized by each subdomain of the protein. As noted above, the difference in these dissociation constants should be at least two orders of magnitude and preferably three or greater.

**(e) Testing for function *in vivo*.**

- 30 Several tests of increasing stringency may be used to confirm the satisfactory performance of a DNA-binding protein designed according to this invention. All share essentially the same components: (1) (a) an expression plasmid directing the production of a chimeric protein comprising the DNA-binding region and a transcriptional activation domain or (b) one or more expression plasmids directing the production of a pair of chimeric proteins of 35 this invention which are capable of dimerizing in the presence of a corresponding dimerizing agent, and thus forming a protein complex containing a DNA-binding region on one protein and a transcription activation domain on the other; and (2) a reporter plasmid directing the

expression of a reporter gene, preferably identical in design to the target gene described above (*i.e.*, multiple binding sites for the DNA-binding domain, a minimal promoter element, and a gene body) but encoding any conveniently measured protein.

In a transient transfection assay, the above-mentioned plasmids are introduced  
5 together into tissue culture cells by any conventional transfection procedure, including for example calcium phosphate co-precipitation, electroporation, and lipofection. After an appropriate time period, usually 24–48 hr, the cells are harvested and assayed for production of the reporter protein. In embodiments requiring dimerization of chimeric proteins for activation of transcription, the assay is conducted in the presence of the dimerizing agent.  
10 In an appropriately designed system, the reporter gene should exhibit little activity above background in the absence of any co-transfected plasmid for the composite transcription factor (or in the absence of dimerizing agent in embodiments under dimerizer control). In contrast, reporter gene expression should be elevated in a dose-dependent fashion by the inclusion of the plasmid encoding the composite transcription factor (or plasmids encoding the  
15 multimerizable chimeras, following addition of multimerizing agent). This result indicates that there are few natural transcription factors in the recipient cell with the potential to recognize the tested binding site and activate transcription and that the engineered DNA-binding domain is capable of binding to this site inside living cells.

The transient transfection assay is not an extremely stringent test in most cases,  
20 because the high concentrations of plasmid DNA in the transfected cells lead to unusually high concentrations of the DNA-binding protein and its recognition site, allowing functional recognition even with relative low affinity interactions. A more stringent test of the system is a transfection that results in the integration of the introduced DNAs at near single-copy. Thus, both the protein concentration and the ratio of specific to non-specific DNA sites would be  
25 very low; only very high affinity interactions would be expected to be productive. This scenario is most readily achieved by stable transfection in which the plasmids are transfected together with another DNA encoding an unrelated selectable marker (*e.g.*, G418-resistance). Transfected cell clones selected for drug resistance typically contain copy numbers of the nonselected plasmids ranging from zero to a few dozen. A set of clones  
30 covering that range of copy numbers can be used to obtain a reasonably clear estimate of the efficiency of the system.

Perhaps the most stringent test involves the use of a viral vector, typically a retrovirus, that incorporates both the reporter gene and the gene encoding the composite transcription factor or multimerizable components thereof. Virus stocks derived from such a  
35 construction will generally lead to single-copy transduction of the genes. If the ultimate application is gene therapy, it may be preferred to construct transgenic animals carrying similar DNAs to determine whether the protein is functional in an animal.

## 10. Design and assembly of the DNA constructs

Constructs may be designed in accordance with the principles, illustrative examples  
5 and materials and methods disclosed in the patent documents and scientific literature cited  
herein, each of which is incorporated herein by reference, with modifications and further  
exemplification as described herein. Components of the constructs can be prepared in  
conventional ways, where the coding sequences and regulatory regions may be isolated, as  
appropriate, ligated, cloned in an appropriate cloning host, analyzed by restriction or  
10 sequencing, or other convenient means. Particularly, using PCR, individual fragments  
including all or portions of a functional unit may be isolated, where one or more mutations may  
be introduced using "primer repair", ligation, in vitro mutagenesis, etc. as appropriate. In the  
case of DNA constructs encoding fusion proteins, DNA sequences encoding individual  
domains and sub-domains are joined such that they constitute a single open reading frame  
15 encoding a fusion protein capable of being translated in cells or cell lysates into a single  
polypeptide harboring all component domains. The DNA construct encoding the fusion  
protein may then be placed into a vector that directs the expression of the protein in the  
appropriate cell type(s). For biochemical analysis of the encoded chimera, it may be  
desirable to construct plasmids that direct the expression of the protein in bacteria or in  
20 reticulocyte-lysate systems. For use in the production of proteins in mammalian cells, the  
protein-encoding sequence is introduced into an expression vector that directs expression in  
these cells. Expression vectors suitable for such uses are well known in the art. Various  
sorts of such vectors are commercially available.

## 25 11. Cells

This invention is particularly useful for the engineering of animal cells and in  
applications involving the use of such engineered animal cells. The animal cells may be  
insect, worm or mammalian or other animal cells. While various mammalian cells may be used,  
including, by way of example, equine, bovine, ovine, canine, feline, murine, and non-human  
30 primate cells, human cells are of particular interest. Among the various species, various types  
of cells may be used, such as hematopoietic, neural, glial, mesenchymal, cutaneous,  
mucosal, stromal, muscle (including smooth muscle cells), spleen, reticuloendothelial,  
epithelial, endothelial, hepatic, kidney, gastrointestinal, pulmonary, fibroblast, and other cell  
types. Of particular interest are hematopoietic cells, which may include any of the nucleated  
35 cells which may be involved with the erythroid, lymphoid or myelomonocytic lineages, as  
well as myoblasts and fibroblasts. Also of interest are stem and progenitor cells, such as



hematopoietic, neural, stromal, muscle, hepatic, pulmonary, gastrointestinal and mesenchymal stem cells

The cells may be autologous cells, syngeneic cells, allogeneic cells and even in some cases, xenogeneic cells with respect to an intended host organism. The cells may be

5 modified by changing the major histocompatibility complex ("MHC") profile, by inactivating beta2-microglobulin to prevent the formation of functional Class I MHC molecules,

inactivation of Class II molecules, providing for expression of one or more MHC molecules, enhancing or inactivating cytotoxic capabilities by enhancing or inhibiting the expression of genes associated with the cytotoxic activity, or the like.

10 In some instances specific clones or oligoclonal cells may be of interest, where the cells have a particular specificity, such as T cells and B cells having a specific antigen specificity or homing target site specificity.

Constructs encoding the chimeric transcription factors or other fusion proteins and constructs comprising target genes can be introduced into the cells as one or more DNA molecules or  
15 constructs, in many cases in association with one or more markers to allow for selection of host cells which contain the construct(s). The construct(s) once completed and demonstrated to have the appropriate sequences may then be introduced into a host cell by any convenient means. The constructs may be incorporated into vectors capable of episomal replication (e.g. BPV or EBV vectors) or into vectors designed for integration into the host  
20 cells' chromosomes. The constructs may be integrated and packaged into non-replicating, defective viral genomes like Adenovirus, Adeno-associated virus (AAV), or Herpes simplex virus (HSV) or others, including retroviral vectors, for infection into cells. Viral delivery systems are discussed in greater detail below. Alternatively, the construct may be introduced by protoplast fusion, electroporation, biolistics, calcium phosphate transfection, lipofection,  
25 microinjection of DNA or the like. The host cells will in some cases be grown and expanded in culture before introduction of the construct(s), followed by the appropriate treatment for introduction of the construct(s) and integration of the construct(s). The cells will then be expanded and screened by virtue of a marker present in the constructs. Various markers which may be used successfully include hpvt, neomycin resistance, thymidine kinase,  
30 hygromycin resistance, etc., and various cell-surface markers such as Tac, CD8, CD3, Thy1 and the NGF receptor.

In some instances, one may have a target site for homologous recombination, where it is desired that a construct be integrated at a particular locus. For example, one can delete and/or replace an endogenous gene (at the same locus or elsewhere) with a recombinant  
35 target construct of this invention. For homologous recombination, one may generally use either  $\Omega$  or O-vectors. See, for example, Thomas and Capecchi, Cell (1987) 51, 503-512; Mansour, et al., Nature (1988) 336, 348-352; and Joyner, et al., Nature (1989) 338, 153-156.

The constructs may be introduced as a single DNA molecule encoding all of the genes, or different DNA molecules having one or more genes. The constructs may be introduced simultaneously or consecutively, each with the same or different markers. Vectors containing useful elements such as bacterial or yeast origins of replication, selectable and/or amplifiable markers, promoter/enhancer elements for expression in prokaryotes or eukaryotes, and mammalian expression control elements, etc. which may be used to prepare stocks of construct DNAs and for carrying out transfections are well known in the art, and many are commercially available.

## 12. Introduction of Constructs into Animals

Cells which have been modified ex vivo with the DNA constructs may be grown in culture under selective conditions and cells which are selected as having the desired construct(s) may then be expanded and further analyzed, using, for example, the polymerase chain reaction for determining the presence of the construct in the host cells and/or assays for the production of the desired gene product(s). Once modified host cells have been identified, they may then be used as planned, e.g. grown in culture or introduced into a host organism.

Depending upon the nature of the cells, the cells may be introduced into a host organism, e.g. a mammal, in a wide variety of ways. Hematopoietic cells may be administered by injection into the vascular system, there being usually at least about  $10^4$  cells and generally not more than about  $10^{10}$  cells. The number of cells which are employed will depend upon a number of circumstances, the purpose for the introduction, the lifetime of the cells, the protocol to be used, for example, the number of administrations, the ability of the cells to multiply, the stability of the therapeutic agent, the physiologic need for the therapeutic agent, and the like. Generally, for myoblasts or fibroblasts for example, the number of cells will be at least about  $10^4$  and not more than about  $10^9$  and may be applied as a dispersion, generally being injected at or near the site of interest. The cells will usually be in a physiologically-acceptable medium.

Cells engineered in accordance with this invention may also be encapsulated, e.g. using conventional biocompatible materials and methods, prior to implantation into the host organism or patient for the production of a therapeutic protein. See e.g. Hguyen et al, Tissue Implant Systems and Methods for Sustaining viable High Cell Densities within a Host, US Patent No. 5,314,471 (Baxter International, Inc.); Uludag and Sefton, 1993, J Biomed. Mater. Res. 27(10):1213-24 (HepG2 cells/hydroxyethyl methacrylate-methyl methacrylate membranes); Chang et al, 1993, Hum Gene Ther 4(4):433-40 (mouse Ltk- cells expressing hGH/immunoprotective perm-selective alginate microcapsules; Reddy et al, 1993, J Infect

Dis 168(4):1082-3 (alginate); Tai and Sun, 1993, FASEB J 7(11):1061-9 (mouse fibroblasts expressing hGH/alginate-poly-L-lysine-alginate membrane); Ao et al, 1995, Transplantation Proc. 27(6):3349, 3350 (alginate); Rajotte et al, 1995, Transplantation Proc. 27(6):3389 (alginate); Lakey et al, 1995, Transplantation Proc. 27(6):3266 (alginate); Korbitt et al, 1995, Transplantation Proc. 27(6):3212 (alginate); Dorian et al, US Patent No. 5,429,821 (alginate); Emerich et al, 1993, Exp Neurol 122(1):37-47 (polymer-encapsulated PC12 cells); Sagen et al, 1993, J Neurosci 13(6):2415-23 (bovine chromaffin cells encapsulated in semipermeable polymer membrane and implanted into rat spinal subarachnoid space); Aebischer et al, 1994, Exp Neurol 126(2):151-8 (polymer-encapsulated rat PC12 cells implanted into monkeys; see also Aebischer, WO 92/19595); Savelkoul et al, 1994, J Immunol Methods 170(2):185-96 (encapsulated hybridomas producing antibodies; encapsulated transfected cell lines expressing various cytokines); Winn et al, 1994, PNAS USA 91(6):2324-8 (engineered BHK cells expressing human nerve growth factor encapsulated in an immunoisolation polymeric device and transplanted into rats); Emerich et al, 1994, Prog Neuropsychopharmacol Biol Psychiatry 18(5):935-46 (polymer-encapsulated PC12 cells implanted into rats); Kordower et al, 1994, PNAS USA 91(23):10898-902 (polymer-encapsulated engineered BHK cells expressing hNGF implanted into monkeys) and Butler et al WO 95/04521 (encapsulated device). The cells may then be introduced in encapsulated form into an animal host, preferably a mammal and more preferably a human subject in need thereof. Preferably the encapsulating material is semipermeable, permitting release into the host of secreted proteins produced by the encapsulated cells. In many embodiments the semipermeable encapsulation renders the encapsulated cells immunologically isolated from the host organism in which the encapsulated cells are introduced. In those embodiments the cells to be encapsulated may express one or more chimeric proteins containing component domains derived from proteins of the host species and/or from viral proteins or proteins from species other than the host species. For example in such cases the chimeras may contain elements derived from GAL4 and VP16. The cells may be derived from one or more individuals other than the recipient and may be derived from a species other than that of the recipient organism or patient.

Instead of *ex vivo* modification of the cells, in many situations one may wish to modify cells *in vivo*. For this purpose, various techniques have been developed for genetic modification of target tissue and cells *in vivo*. A number of viral vectors have been developed, such as adenovirus, adeno-associated virus, and retroviruses, which allow for transduction and, in some cases, integration of the virus into the host. See, for example, Dubensky et al. (1984) Proc. Natl. Acad. Sci. USA 81, 7529-7533; Kaneda et al., (1989) Science 243,375-378; Hiebert et al. (1989) Proc. Natl. Acad. Sci. USA 86, 3594-3598; Hatzoglu et al. (1990) J. Biol. Chem. 265, 17285-17293 and Ferry, et al. (1991) Proc. Natl.

Acad. Sci. USA 88, 8377-8381. The vector may be administered by injection, e.g. intravascularly or intramuscularly, inhalation, or other parenteral mode. Non-viral delivery methods such as administration of the DNA via complexes with liposomes or by injection, catheter or biolistics may also be used. See e.g. WO 96/41865, PCT/US97/22454 and  
5 USSN 60/084819, for example, for additional guidance on formulation and delivery of recombinant nucleic acids to cells and to organisms. Those references as well as the references cited previously, including those relating to tetR-based systems, progesterone-r-based systems and ecdysone-based systems provide detailed additional guidance on the preparation, formulation and delivery of various ligands to cells in vitro and to organisms. As  
10 mentioned elsewhere, the contents of those cited documents are incorporated herein by reference.

In accordance with *in vivo* genetic modification, the manner of the modification will depend on the nature of the tissue, the efficiency of cellular modification required, the number of opportunities to modify the particular cells, the accessibility of the tissue to the DNA  
15 composition to be introduced, and the like. By employing an attenuated or modified retrovirus carrying a target transcriptional initiation region, if desired, one can activate the virus using one of the subject transcription factor constructs, so that the virus may be produced and transfect adjacent cells.

The DNA introduction need not result in integration in every case. In some situations, transient maintenance of the DNA introduced may be sufficient. In this way, one could have  
20 a short term effect, where cells could be introduced into the host and then turned on after a predetermined time, for example, after the cells have been able to home to a particular site.

### 13. Applications

This invention is applicable to any situation that calls for expression of an endogenous or exogenously-introduced gene, e.g. that is embedded within a large genome. The desired expression level could be preset very high or very low. The system may be further engineered to achieve regulated or titratable expression. See e.g. PCT/US93/01617 and other previously cited references. In most cases, the inadvertent activation of unrelated  
25 30 cellular genes is undesirable.

**1. Regulated high-level gene expression in gene therapy.** Gene therapy often requires controlled high-level expression of a therapeutic gene, sometimes in a cell-type specific pattern. By supplying the therapeutic gene with saturating amounts of an activating  
35 transcription factor in accordance with this invention, considerably higher levels of gene expression can be obtained relative to natural promoters or enhancers, which are dependent on endogenous transcription factors. In addition, by supplying a ligand binding domain,

regulated expression of the therapeutic gene may be achieved. Thus, one application of this invention to gene therapy is the delivery of a two-transcription-unit cassette (which may reside on one or two plasmid molecules, depending on the delivery vector) consisting of (1) a transcription unit encoding a transcription factor, whether naturally occurring or designed as described above, for instance comprising a composite DNA-binding domain and a strong transcription activation domain (e.g., derived from the VP16 protein or a human transcription factor) linked to a ligand binding domain which, upon binding to ligand regulates transcription of the target gene and (2) a transcription unit consisting of the target gene linked to and under the control of a minimal promoter carrying one, and preferably several, binding sites for the composite DNA-binding domain. Cointroduction of the two transcription units into a cell results in the production of the hybrid transcription factor which in turn activates the therapeutic gene to high level. This strategy essentially incorporates an amplification step, because the promoter that would be used to produce the therapeutic gene product in conventional gene therapy is used instead to produce the activating transcription factor. Each transcription factor has the potential to direct the production of multiple copies of the therapeutic protein. In some cases, the application may involve a three-transcription unit cassette, where the DNA binding domain and activation domains are each linked to a ligand binding domain. In such cases, binding of the ligand to each of the ligand binding domains reconstitutes the transcription factor and allows transcription to proceed, as described in US 5,830,462, the contents of which are incorporated herein by reference.

This method may be employed to increase the efficacy of many gene therapy strategies by substantially elevating the expression of a therapeutic target gene, allowing expression to reach therapeutically effective levels. Examples of therapeutic genes that would benefit from this strategy are genes that encode secreted therapeutic proteins, such as cytokines (e.g., IL-2, IL-4, IL-12), CFTR (see e.g. Grubb et al, 1994, Nature 371:802-6), growth factors (e.g., VEGF), antibodies, and soluble receptors. Other candidate therapeutic genes are disclosed in PCT/US93/01617. This strategy may also be used to increase the efficacy of "intracellular immunization" agents, molecules like ribozymes, antisense RNA, and dominant-negative proteins, that act either stoichiometrically or by competition. Examples include agents that block infection by or production of HIV or hepatitis virus and agents that antagonize the production of oncogenic proteins in tumors.

It should be appreciated that in practice, the system is subject to many variables, such as the efficiency of expression and, as appropriate, the level of secretion, the activity of the expression product, the particular need of the patient, which may vary with time and circumstances, the rate of loss of the cellular activity as a result of loss of cells or expression activity of individual cells, and the like. Therefore, it is expected that for each individual

patient, even if there were universal cells which could be administered to the population at large, each patient would be monitored for the proper dosage for the individual.

**2. Production of recombinant proteins.** Production of recombinant therapeutic proteins for

commercial and investigational purposes is often achieved through the use of mammalian cell lines engineered to express the protein at high level. The use of mammalian cells, rather than bacteria or yeast, is indicated where the proper function of the protein requires post-translational modifications not generally performed by heterologous cells. Examples of proteins produced commercially this way include erythropoietin, tissue plasminogen activator, clotting factors such as Factor VIII:c, antibodies, etc. The cost of producing proteins in this fashion is directly related to the level of expression achieved in the engineered cells. Thus, because the constitutive two-transcription-unit system described above can achieve considerably higher expression levels than conventional expression systems, it may greatly reduce the cost of protein production.

**3. Biological research.** This invention is applicable to a wide range of biological experiments in which precise control over a target gene is desired. These include: (1) expression of a protein or RNA of interest for biochemical purification; (2) tissue or organ specific expression of a protein or RNA of interest in transgenic animals for the purposes of evaluating its biological function. Transgenic animal models and other applications for which this invention may be used include those disclosed in US Patent Application Serial Nos. 08/292,595 and 08/292,596 (filed August 18, 1994).

This invention further provides kits useful for practicing the described methods. Such kits contain a first DNA sequence encoding a transcription factor and a second DNA sequence containing a target gene linked to a DNA element to which the transcription factor is capable of binding. Alternatively, the second DNA sequence may contain a cloning site for insertion of a desired target gene by the practitioner. The kits optionally also contain a ligand useful for regulated expression of the target gene.

~~~~~

The following examples contain important additional information, exemplification and guidance which can be adapted to the practice of this invention in its various embodiments and the equivalents thereof. The examples are offered by way illustration should not be construed as limiting in any way. The contents of all cited references including literature references, issued patents, published patent applications as cited throughout this application

are hereby expressly incorporated by reference. The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See,

5 for example, Molecular Cloning†A Laboratory†Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal

10 Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.), Immunochemical Methods In Cell And Molecular Biology

15 (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

## Examples

I. Individual DNA-binding and transcription activating components are modular, may be incorporated into fusion proteins with various other domains and function as intended in cell culture and in animals:

A. ZFHD1 and p65 work well individually in cell culture and in whole animals in drug-dependent (regulatable) transcription systems

### 1. Constructs encoding chimeric transcription factors

(a) Unless otherwise stated, all DNA manipulations described in this and other examples were performed using standard procedures (See e.g., F.M. Ausubel *et al.*, Eds., Current Protocols in Molecular Biology (John Wiley & Sons, New York, 1994).

#### (b) Plasmids

Constructs encoding fusions of human FKBP12 (hereafter 'FKBP') with the yeast GAL4 DNA binding domain, the HSV VP16 activation domain, human T cell CD3 zeta chain intracellular domain or the intracellular domain of human FAS are disclosed in PCT/US94/01617.

Additional DNA vectors for directing the expression of fusion proteins relevant to this invention were derived from the mammalian expression vector pCGNN (Attar, R.M. and Gilman, M.Z. 1992. *MCB* 12: 2432-2443). Inserts cloned as XbaI-BamHI fragments into pCGNN are transcribed under the control of the human CMV promoter and enhancer sequences (nucleotides -522 to +72 relative to the cap site), and are expressed with an optional epitope tag (a 16 amino acid portion of the *H. influenzae* hemagglutinin gene that is recognized by the monoclonal antibody 12CA5) and, in the case of transcription factor domains, with an N-terminal nuclear localization sequence (NLS; from SV40 T antigen). Except where stated, all fragments cloned into pCGNN were inserted as XbaI-BamHI fragments that included a SpeI site just upstream of the BamHI site. As XbaI and SpeI produce compatible ends, this allowed further XbaI-BamHI fragments to be inserted downstream of the initial insert and facilitated stepwise assembly of proteins comprising multiple components. A stop codon was interposed between the SpeI and BamHI sites. For initial constructs, the vector pCGNN-GAL4 was additionally used, in which codons 1-94 of the GAL4 DNA binding domain gene were cloned into the XbaI site of pCGNN such that a XbaI site is regenerated only at the 3' end of the fragment. Thus XbaI-BamHI fragments could be cloned into this vector to generate GAL4 fusions, and subsequently recovered.



(c) Constructs encoding GAL4 DNA binding domain- FRAP fusions

To obtain portions of the human FRAP gene, human thymus total RNA (Clontech #64028-1) was reverse transcribed using MMLV reverse transcriptase and random hexamer primer (Clontech 1st strand synthesis kit). This cDNA was used directly in a PCR reaction

5 containing primers **1** and **2** and Pfu polymerase (Stratagene). The primers were designed to amplify the coding sequence for amino acids 2025-2113 inclusive of human FRAP: an 89 amino acid region essentially corresponding to the minimal 'FRB' domain identified by Chen et al. (*Proc. Natl. Acad. Sci. USA* (1995) 92, 4947-4951) as necessary and sufficient for FKBP-rapamycin binding (hereafter named FRB). The appropriately-sized band was purified,  
10 digested with XbaI and SpeI, and ligated into XbaI-SpeI digested pCGNN-GAL4. This construct was confirmed by restriction analysis (to verify the correct orientation) and DNA sequencing and designated pCGNN-GAL4-1FRB.

Constructs encoding FRB multimers were obtained by isolating the FRB XbaI-BamHI fragment, and then ligating it back into pCGNN-GAL4-1FRB digested with SpeI and BamHI  
15 to generate pCGNN-GAL4-2FRB, which was confirmed by restriction analysis. This procedure was repeated analogously on the new construct to yield pCGNN-GAL4-3FRB and pCGNN-GAL4-4FRB.

Vectors were also constructed that encode larger fragments of FRAP, encompassing the minimal FRB domain (amino acids 2025-2113) but extending beyond it. PCR primers were  
20 designed that amplify various regions of FRAP flanked by 5' XbaI and 3' SpeI sites as indicated below.

|    | <b><i>Designation</i></b> | <b><i>amino acids</i></b> | <b><i>5' primer</i></b> | <b><i>3' primer</i></b> |
|----|---------------------------|---------------------------|-------------------------|-------------------------|
|    | FRAP <sub>a</sub>         | 2012-2127                 | <b>6</b>                | <b>7</b>                |
| 25 | FRAP <sub>b</sub>         | 1995-2141                 | <b>5</b>                | <b>8</b>                |
|    | FRAP <sub>c</sub>         | 1945-2113                 | <b>3</b>                | <b>2</b>                |
|    | FRAP <sub>d</sub>         | 1995-2113                 | <b>5</b>                | <b>2</b>                |
|    | FRAP <sub>e</sub>         | 2012-2113                 | <b>6</b>                | <b>2</b>                |
|    | FRAP <sub>f</sub>         | 2025-2127                 | <b>1</b>                | <b>7</b>                |
| 30 | FRAP <sub>g</sub>         | 2025-2141                 | <b>1</b>                | <b>8</b>                |
|    | FRAP <sub>h</sub>         | 2025-2174                 | <b>1</b>                | <b>4</b>                |
|    | FRAP <sub>i</sub>         | 1945-2174                 | <b>3</b>                | <b>4</b>                |

Initially, fragment FRAP<sub>i</sub> was amplified by RT-PCR as described above, digested with XbaI  
35 and SpeI, and ligated into XbaI-SpeI digested pCGNN-GAL4. This construct, pCGNN-

GAL4-FRAP<sub>i</sub>, was analyzed by PCR to confirm insert orientation and verified by DNA sequencing. It was then used as a PCR substrate to amplify the other fragments using the primers listed. The new fragments were cloned as GAL4 fusions as described above to yield the constructs pCGNN-GAL4-FRAP<sub>a</sub>, pCGNN-GAL4-FRAP<sub>b</sub> etc, which were confirmed by

5 DNA sequencing.

Vectors encoding concatenates of two of the larger FRAP fragments, FRAP<sub>d</sub> and FRAP<sub>e</sub>, were generated by analogous methods to those used earlier. XbaI-BamHI fragments encoding FRAP<sub>d</sub> and FRAP<sub>e</sub> were isolated from pCGNN-GAL4-FRAP<sub>d</sub> and pCGNN-GAL4-FRAP<sub>e</sub> and ligated back into the same vectors digested with SpeI and BamHI to generate

10 pCGNN-GAL4-2FRAP<sub>d</sub> and pCGNN-GAL4-2FRAP<sub>e</sub>. This procedure was repeated analogously on the new constructs to yield pCGNN-GAL4-3FRAP<sub>d</sub>, pCGNN-GAL4-3FRAP<sub>e</sub>, pCGNN-GAL4-4FRAP<sub>d</sub> and pCGNN-GAL4-4FRAP<sub>e</sub>. All constructs were verified by restriction analysis.

15 (d) Constructs encoding FRAP-VP16 activation domain fusions

To generate N-terminal fusions of FRB domain(s) with the activation domain of the Herpes Simplex Virus protein VP16, the XbaI-BamHI fragments encoding 1, 2, 3 and 4 copies of FRB were recovered from the GAL4 fusion vectors and ligated into XbaI-BamHI digested pCGNN to yield pCGNN-1FRB, pCGNN-2FRB etc. These vectors were then digested with SpeI

20 and BamHI. An XbaI-BamHI fragment encoding amino acids 414-490 of VP16 was isolated from plasmid pCG-Gal4-VP16 (Das, G., Hinkley, C.S. and Herr, W. (1995) *Nature* 374, 657-660) and ligated into the SpeI-BamHI digested vectors to generate pCGNN-1FRB-VP16, pCGNN-2FRB-VP16, etc. The constructs were verified by restriction analysis and/or DNA sequencing.

25

(e) Constructs encoding ZFHD1 DNA binding domain- FRAP fusions

An expression vector for directing the expression of ZFHD1 coding sequence in mammalian cells was prepared as follows. Zif268 sequences were amplified from a cDNA clone by PCR using primers 5'Xba/Zif and 3'Zif+G. Oct1 homeodomain sequences were amplified from a

30 cDNA clone by PCR using primers 5'Not Oct HD and Spe/Bam 3'Oct. The Zif268 PCR fragment was cut with XbaI and NotI. The Oct1 PCR fragment was cut with NotI and BamHI. Both fragments were ligated in a 3-way ligation between the XbaI and BamHI sites of pCGNN (Attar and Gilman, 1992) to make pCGNNZFHD1 in which the cDNA insert is under the transcriptional control of human CMV promoter and enhancer sequences and is linked to

35 the nuclear localization sequence from SV40 T antigen. The plasmid pCGNN also contains a gene for ampicillin resistance which can serve as a selectable marker. (Derivatives,

pCGNNZFHD1-FKBPx1 and pCGNNZFHD1-FKBPx3, were prepared containing one or three tandem repeats of human FKBP12 ligated as an XbaI-BamHI fragment between the SpeI and BamHI sites of pCGNNZFHD1. A sample of pCGNNZFHD1-FKBPx3 has been deposited with the American Type Culture Collection under ATCC Accession No. 97399.)

5

# **Primers:**

5'Xba/Zif                      5'ATGCTCTAGAGAACGCCCATATGCTTGCCCT  
3'Zif+G                      5'ATGCGCGGCCGCGCCTGTGTGGGTGCGGATGTG

10 5'Not OctHD    5'ATGCGCGGCCGCGCAGGAGGAAGAAACGCACCAGC  
Spe/Bam 3'Oct 5'GCATGGATCCGATTCAACTAGTGTTGATTCTTTTTCTTTCTGGCGGCG

To generate C-terminal fusions of FRB domain(s) with the chimeric DNA binding protein ZFHD1, the XbaI-BamHI fragments encoding 1, 2, 3 and 4 copies of FRB were recovered from the GAL4 fusion vectors and ligated into Spe-BamHI digested pCGNN-ZFHD1 to yield pCGNN-ZFHD1-1FRB, pCGNN-ZFHD1-2FRB etc. Constructs were verified by restriction analysis and/or DNA sequencing.

To examine the effect of introducing additional 'linker' polypeptide between ZFHD1 and a C-terminal FRB domain, FRAP fragments encoding extra sequence N-terminal to FRB were cloned as ZFHD1 fusions. XbaI-BamHI fragments encoding FRAP<sub>a</sub>, FRAP<sub>b</sub>, FRAP<sub>c</sub>, FRAP<sub>d</sub> and FRAP<sub>e</sub> were excised from the vectors pCGNN-GAL4-FRAP<sub>a</sub>, pCGNN-GAL4-FRAP<sub>b</sub> etc and ligated into SpeI-BamHI digested pCGNN-ZFHD1 to yield the vectors pCGNN-ZFHD1-FRAP<sub>a</sub>, pCGNN-ZFHD1-FRAP<sub>b</sub>, etc. Vectors encoding fusions of ZFHD1 to 2, 3 and 4 C-terminal copies of FRAP<sub>e</sub> were also constructed by isolating XbaI-BamHI fragments encoding 2FRAP<sub>e</sub>, 3FRAP<sub>e</sub> and 4FRAP<sub>e</sub> from pCGNN-GAL4-2FRAP<sub>e</sub>, pCGNN-GAL4-3FRAP<sub>e</sub> and pCGNN-GAL4-4FRAP<sub>e</sub> and ligating them into SpeI-BamHI digested pCGNN-ZFHD1 to yield the vectors pCGNN-ZFHD1-2FRAP<sub>e</sub>, pCGNN-ZFHD1-3FRAP<sub>e</sub> and pCGNN-ZFHD1-4FRAP<sub>e</sub>. All constructs were verified by restriction analysis.

Vectors were also constructed that encode N-terminal fusions of FRB domain(s) with ZFHD1. XbaI-BamHI fragments encoding 1, 2, 3 and 4 copies of FRAP<sub>e</sub> were isolated from pCGNN-GAL4-1FRAP<sub>e</sub>, pCGNN-GAL4-2FRAP<sub>e</sub> etc and ligated into XbaI-BamHI digested pCGNN to yield the plasmids pCGNN-1FRAP<sub>e</sub>, pCGNN-2FRAP<sub>e</sub> etc. These vectors were then digested with SpeI and BamHI, and an XbaI-BamHI fragment encoding ZFHD1 (isolated from pCGNN-ZFHD1) ligated in to yield the constructs pCGNN-1FRAP<sub>e</sub>-ZFHD1, pCGNN-2FRAP<sub>e</sub>-ZFHD1 etc, which were verified by restriction analysis.

(f) Constructs encoding FRAP-p65 activation domain fusions

To generate fusions of FRB domain(s) with the activation domain of the human NF-kb p65 subunit (hereafter designated p65), two fragments were amplified by PCR from the plasmid pCG-p65. Primers **9** (p65/ 5' Xba) and **11** (p65 3' Spe/Bam) amplify the coding sequence for amino acids 450-550, and primers **10** (p65/361 Xba) and **11** amplify the coding sequence for amino acids 361-550, both flanked by 5' XbaI and 3' SpeI/BamHI sites. PCR products were digested with XbaI and BamHI and cloned into XbaI-BamHI digested pCGNN to yield pCGNN-p65(450-550) and pCGNN-p65(361-550). The constructs were verified by restriction analysis and DNA sequencing.

The 100 amino acid P65 transcription activation sequence is encoded by the following linear sequence:

```
CTGGGGGCCTTGCTTGGCAACAGCACAGACCCAGCTGTGTTACAGACCTGGCATCCGTCGACAA
CTCCGAGTTTCAGCAGCTGCTGAACCAGGGCATACTGTGGCCCCCACACAACTGAGCCCATGC
TGATGGAGTACCTGAGGCTATAACTCGCCTAGTGACAGGGGGCCAGAGGCCCCCGACCCAGCT
CCTGCTCCACTGGGGGCCCCGGGGCTCCCCAATGGCCTCCTTTCAGGAGATGAAGACTTCTCCTC
CATTGCGGACATGGACTTCTCAGCCCTGCTGAGTCAGATCAGCTCC
```

The more extended p65 transcription activation domain (351-550) is encoded by the following linear sequence:

```
GATGAGTTTCCCACCATGGTGTTCCTTCTGGGCAGATCAGCCAGGCCTCGGCCTTGGCCCCGGCC
CCTCCCCAAGTCCTGCCCCAGGCTCCAGCCCCTGCCCTGCTCCAGCCATGGTATCAGCTCTGGC
CCAGGCCCCAGCCCCTGTCCCAGTCCTAGCCCCAGGCCCTCCTCAGGCTGTGGCCCCACCTGCCC
CCAAGCCCACCCAGGCTGGGGAAGGAACGCTGTGACAGGGCCCTGCTGCAGCTGCAGTTTGATGAT
GAAGACCTGGGGGCCTTGCTTGGCAACAGCACAGACCCAGCTGTGTTACAGACCTGGCATCCGT
CGACAACTCCGAGTTTCAGCAGCTGCTGAACCAGGGCATACTGTGGCCCCCACACAACTGAGC
CCATGCTGATGGAGTACCCTGAGGCTATAACTCGCCTAGTGACAGCCCAGAGGCCCCCGACCCA
GCTCCTGCTCCACTGGGGGCCCCGGGGCTCCCCAATGGCCTCCTTTCAGGAGATGAAGACTTCTC
CTCCATTGCGGACATGGACTTCTCAGCCCTGCTGAGTCAGATCAGCTCCTAA
```

To generate N-terminal fusions of FRB domain(s) with portions of the p65 activation domain, plasmids pCGNN-1FRB, pCGNN-2FRB etc were digested with SpeI and BamHI. An XbaI-BamHI fragment encoding p65 (450-550) was isolated from pCGNN-p65(450-550) and ligated into the SpeI-BamHI digested vectors to yield the plasmids pCGNN-1FRB-p65(450-

550), pCGNN-2FRB-p65(450-550) etc. The construct pCGNN-1FRB-p65(361-550) was made similarly using an XbaI-BamHI fragment isolated from pCGNN-p65(361-550). These constructs were verified by restriction analysis.

To examine the effect of introducing additional 'linker' polypeptide between the p65 activation domain and an N-terminal FRB domain, FRAP fragments encoding extra sequence C-terminal to FRB were cloned as p65 fusions. XbaI-BamHI fragments encoding FRAP<sub>a</sub>, FRAP<sub>b</sub>, FRAP<sub>f</sub>, FRAP<sub>g</sub> and FRAP<sub>h</sub> were excised from the vectors pCGNN-GAL4-FRAP<sub>a</sub>, pCGNN-GAL4-FRAP<sub>b</sub> etc and ligated into XbaI-BamHI digested pCGNN to yield the vectors pCGNN-FRAP<sub>a</sub>, pCGNN-FRAP<sub>b</sub>, etc. These plasmids were then digested with SpeI and BamHI, and a XbaI-BamHI fragment encoding p65 (amino acids 450-550) ligated in to yield the five vectors pCGNN-FRAP<sub>a</sub>-p65, pCGNN-FRAP<sub>b</sub>-p65, etc, which were verified by restriction analysis.

Vectors encoding fusions of p65 to 1 and 3 N-terminal copies of FRAP<sub>e</sub> were also prepared by digesting pCGNN-1FRAP<sub>e</sub> and pCGNN-3FRAP<sub>e</sub> with SpeI and BamHI. XbaI-BamHI fragments encoding p65(450-550) and p65(361-550) (isolated from pCGNN-p65(450-550) and pCGNN-p65(361-550)) were then ligated in to yield the vectors pCGNN-1FRAP<sub>e</sub>-p65(450-550), pCGNN-3FRAP<sub>e</sub>-p65(450-550), pCGNN-1FRAP<sub>e</sub>-p65(361-550) and pCGNN-3FRAP<sub>e</sub>-p65(361-550). All constructs were verified by restriction analysis.

Vectors were also constructed that encode C-terminal fusions of FRB domain(s) with portions of the p65 activation domain. Plasmids pCGNN-p65(450-550) and pCGNN-p65(361-550) were digested with SpeI and BamHI, and XbaI-BamHI fragments encoding 1 and 3 copies of FRAP<sub>e</sub> (isolated from pCGNN-GAL4-1FRAP<sub>e</sub> and pCGNN-GAL4-3FRAP<sub>e</sub>) and 1 copy of FRB (isolated from pCGNN-GAL4-1FRB) ligated in to yield the plasmids pCGNN-p65(450-550)-1FRAP<sub>e</sub>, pCGNN-p65(450-550)-3FRAP<sub>e</sub>, pCGNN-p65(361-550)-1FRAP<sub>e</sub>, pCGNN-p65(361-550)-3FRAP<sub>e</sub>, pCGNN-p65(450-550)-1FRB and pCGNN-p65(361-550)-1FRB. All constructs were verified by restriction analysis.

#### (g) Further constructs

Other constructs can be made analogously with the above procedures, but using alternative portions of the FRAP sequence. For example, primers **12** and **13** are used to amplify the entire coding region of FRAP. Primers **1** and **13**, **6** and **13**, and **5** and **13**, are used to amplify three fragments encompassing the FRB domain and extending through to the C-terminal end of the protein (including the lipid kinase homology domain). These fragments differ by encoding different portions of the protein N-terminal to the FRB domain. In each case, RT-PCR is used as described above to amplify the regions from human thymus RNA, the PCR

products are purified, digested with XbaI and SpeI, ligated into XbaI-SpeI digested pCGNN, and verified by restriction analysis and DNA sequencing.

(h) Primer sequences

```
5 1 5' GCATGTCTAGAGAGATGTGGCATGAAGGCCTGGAAG
2 5' GCATCACTAGTCTTTGAGATTCGTCTGGAACACATG
3 5' GCACATTCTAGAAATTGATACGCCCAGACCCTTG
4 5' CGATCAACTAGTAAGTGTCAATTTCCGGGGCCT
5 5' GCACTATCTAGACTGAAGAACATGTGTGAGCACAGC
10 6 5' GCACTATCTAGAGTGAGCGAGGAGCTGATCCGAGTG
7 5' CGATCAACTAGTGGAAACATATTGCAGCTCTAAGGA
8 5' CGATCAACTAGTTGGCACAGCCAATTCAAGGTCCCG
9 5' ATGCTCTAGACTGGGGGCCTTGCTTGGCAAC
10 5' ATGCTCTAGAGATGAGTTTCCCACCATGGTG
15 11 5' GCATGGATCCGCTCAACTAGTGGAGCTGATCTGACTCAG
12 5' ATGCTCTAGACTTGAACCGGACCTGCCGCC
13 5' GCATCACTAGTCCAGAAAGGGCACCAGCCAATAT
```

Restriction sites are underlined (XbaI = TCTAGA, SpeI = ACGAGT, BamHI = GGATCC).

20

(i) DNA sequence of representative final construct: pCGNN-ZFHD1-1FRB

12CA5

5

epitope

M A S S Y P Y D V P D  
5' gtagaagcgcggt ATG GCT TCT AGC TAT CCT TAT GAC GTG CCT GAC

10

SV40 T NLS

Y A S L G G P S S P K K K R K  
TAT GCC AGC CTG GGA GGA CCT TCT AGT CCT AAG AAG AAG AGA AAG  
(X/S)

15

ZFHD1 (5')

V S R E R P Y A C P V E S C D...  
GTG TCT AGA GAA CGC CCA TAT GCT TGC CCT GTC GAG TCC TGC GA...

20

XbaI

ZFHD1 (3')

FRB (5')

... R I N T R E M W H E G L E E...  
25 ...AGA ATC AAC ACT AGA GAG ATG TGG CAT GAA GGC CTG GAA GA...  
(S/X)

FRB (3')

R I S K T S Y \*  
30 CGA ATC TCA AAG ACT AGT TAT TAG ggatcctgag  
SpeI BamHI

Non-coding nucleotides are indicated in lower case

- 35 (S/X) and (X/S) indicate the result of a ligation event between the compatible products of digestion with XbaI and SpeI, to produce a sequence that is cleavable by neither enzyme  
\* indicates a stop codon

(j) Bicistronic constructs

The internal ribosome entry sequence (IRES) from the encephalomyocarditis virus was amplified by PCR from pWZL-Bleo. The resulting fragment, which was cloned into pBS-SK+ (Stratagene), contains an XbaI site and a stop codon upstream of the IRES sequence and downstream of it, an NcoI site encompassing the ATG followed by SpeI and BamHI sites. To facilitate cloning, the sequence around the initiating ATG of pCGNN-ZFHD1-3FKBP was mutated to an NcoI site and the XbaI site was mutated to a NheI site using the oligonucleotides

5'-GAATTCCTAGAAGCGACCATGGCTTCTAGC-3'

and

5'-GAAGAGAAAGGTGGCTAGCGAACGCCCATAT-3'

respectively. An NcoI-BamHI fragment containing ZFHD1-3FKBP was then cloned downstream of pBS-IRES to create pBS-IRES-ZFHD1-3FKBP. The XbaI-BamHI fragment from this plasmid was next cloned into SpeI/BamHI-cut pCGNN-1FRB-p65(361-550) to create pCGNN-1FRB-p65(361-550)-IRES-ZFHD1-3FKBP.

PCR primers for the HSF trimerization domain:

5'primer: atgctctagaagtgtgtccaccctgaagagtgaagac

3' primer: atgctgatcaagatctttattaactagtgccactgtcgttcagcatcagggggat

Template: pBS108 vector containing human HSF1 full length cDNA.

The PCR fragments containing the trimerization domain of HSF1 (amino acids 120-217) will be digested with XbaI/BclI and cloned into SpeI/BamHI digested PCGNN-FRB vectors. The resulting PCGNN-FRB-T vectors will be digested with SpeI-Bgl II and ligated with fragments containing p65 activation domain (amino acids 361-550) or constructs containing both the p65 activation domain and the HSF activation domain (amino acids 431-529).

## **2. Retroviral vectors for the expression of chimeric proteins**

Retroviral vectors used to express transcription factor fusion proteins from stably integrated, low copy genes were derived from pSRaMSVtkNeo (Muller et al., MCB 11:1785-92, 1991) and pSRaMSV(XbaI) (Sawyers et al., J. Exp. Med. 181:307-313, 1995). Unique BamHI sites in both vectors were removed by digesting with BamHI, filling in with Klenow and religating to produce pSMTN2 and pSMTX2, respectively. pSMTN2 expresses the Neo



gene from an internal thymidine kinase promoter. A Zeocin gene (Invitrogen) will be cloned as a NheI fragment into a unique XbaI site downstream of an internal thymidine kinase promoter in pSMTX2 to yield pSNTZ. This Zeocin fragment was generated by mutagenizing pZeo/SV (Invitrogen) using the following primers to introduce NheI sites flanking the zeocin coding sequence.

Primer 1 5'-GCCATGGTGGCTAGCCTATAGTGAG  
Primer2 5'-GGCGGTGTTGGCTAGCGTCGGTCAG

pSMTN2 contains unique EcoRI and HindIII sites downstream of the LTR. To facilitate cloning of transcription factor fusion proteins synthesized as XbaI-BamHI fragments the following sequence was inserted between the EcoRI and HindIII sites to create pSMTN3:

12CA5 epitope

M A S S Y P Y D V P D  
5' gaattccagaagcgct ATG GCT TCT AGC TAT CCT TAT GAC GTG CCT GAC  
EcoRI

SV40 T NLS

Y A S L G G P S S P K K K R K  
TAT GCC AGC CTG GGA GGA CCT TCT AGT CCT AAG AAG AAG AGA AAG

V  
GTG TCT AGA TAT CGA GGA TCC CAA GCT T  
XbaI BamHI HindIII

The equivalent fragment is inserted into a unique EcoRI site of pSMTZ to create pSMTZ3 with the only difference being that the 3' HindIII site is replaced by an EcoRI site. pSMTN3 and pSMTZ3 permit chimeric transcription factors to be cloned downstream of the 5' viral LTR as XbaI-BamHI fragments and allow selection for stable integrants by virtue of their ability to confer resistance to the antibiotics G418 or Zeocin respectively. To generate the retroviral vector SMTN-ZFHD1-3FKBP, pCGNN-ZFHD1-3FKBP was first mutated to add an EcoRI site upstream of the first amino acid of the fusion protein. An EcoRI-BamHI(blunted) fragment was then cloned into EcoRI-HindIII(blunted) pSRaMSVtkNeo (ref. 51) so that ZFHD1-3FKBP was expressed from the retroviral LTR.

### 3. Rapamycin-dependent transcriptional activation

Our previous experiments showed that three copies of FKBP fused either to a Gal4 DNA binding domain or a transcription activation domain activated both the stably integrated or transiently transfected reporter gene more strongly than corresponding fusion proteins containing only one or two FKBP domains. To evaluate this parameter with FRB fusion proteins, effector plasmids containing Gal4 DNA binding domain fused to one or more copies of an FRB domain were co-transfected with a plasmid encoding three FKBP domains and a p65 activation domain (3xFKBP-p65) by transient transfection. The results indicate that in this system, four copies of the FRB domain fused to the Gal4 DNA binding domain activated the stably integrated reporter gene more strongly than other corresponding fusion proteins with fewer FRB domains.

**Method:** HT1080 B cells were grown in MEM supplemented with 10% Bovine Calf Serum. Approximately  $4 \times 10^5$  cells/well in a 6 well plate (Falcon) were transiently transfected by Lipofection procedure as recommended by the supplier (GIBCO, BRL). The DNA: Lipofectamine ratio used in this experiment correspond to 1:6. Cells in each well recieved 500 ng of pCGNN F3-p65, 1.9 ug of PUC 118 plasmid as carrier and 100 ng of one of the following plasmids: pCGNN Gal4-1FRB, pCGNN Gal4-2FRB, pCGNN Gal4-3FRB or pCGNN Gal4-4FRB. Following transfection, 2 ml fresh media was added and supplemented with Rapamycin to the indicated concentration. After 24 hrs, 100 ul of the media was assayed for SEAP activity as described (Spencer et al, 1993).

To test whether multiple FRB domains fused to a p65 activation domain results in increased transcriptional activation of the reporter gene, we co-transfected HT1080 B cells with plasmids expressing Gal4-3xFKBP and 1, 2, 3 or 4 copies of FRB fused to p65 activation domain. Surprisingly, unlike the DNA binding domain-FRB fusions, a single copy of FRB fused to p65 activation domain activated the reporter gene significantly more strongly than corresponding fusion proteins containing 2 or more copies of FRB.

**Method:** HT1080 B cells were grown in MEM supplemented with 10% Bovine Calf Serum. Approximately  $4 \times 10^5$  cells/well in a 6 well plate were transiently transfected by Lipofection procedure as recommended by GIBCO, BRL. The DNA: Lipofectamine ratio used correspond to 1:6. Cells in each well recieved 1.9 ug of PUC 118 plasmid as carrier , 100 ng of pCGNNGal4F3 and 500 ng one of the following plasmids :pCGNN1, 2, 3 or 4 FRB-p65. Following transfection, 2 ml fresh media was added and supplemented with Rapamycin to

the indicated concentration. After 24 hrs, 100 ul of the media was assayed for SEAP activity as described (Spencer et al, 1993).

Similar experiments were also conducted using another stable cell line (HT1080 B14) containing the 5xGal4-IL2-SEAP reporter gene and DNA sequences encoding a fusion

- 5 protein containing a Gal4 DNA binding domain and 3 copies of FKBP stably integrated. These cells were transiently transfected with effector plasmids expressing p65 activation domain fused to 1 or more copies of an FRB domain. Similar to our observations with HT1080 B cells, effector plasmids expressing a single copy of FRB-p65 activation domain fusion protein activated the reporter gene more strongly than others with 2 or more copies of FRB.

10

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

#### 4. Rapamycin-dependent transcriptional activation in transiently transfected cells: ZFHD1 and p65 fusions

Human fibrosarcoma cells transiently transfected with a SEAP target gene and plasmids encoding representative ZFHD-FKBP- and FRB-p65-containing fusion proteins exhibited rapamycin-dependent and dose-responsive secretion of SEAP into the cell culture medium. SEAP production was not detected in cells in which one or both of the transcription factor fusion plasmids was omitted, nor was it detected in the absence of added rapamycin. When all components were present, however, SEAP secretion was detectable at rapamycin concentrations as low as 0.5 nM. Peak SEAP secretion was observed at 5 nM. Similar results have been obtained when the same transcription factors were used to drive rapamycin-dependent activation of an hGH reporter gene or a stably integrated version of the SEAP reporter gene made by infection with a retroviral vector. It is difficult to determine the fold activation in response to rapamycin since levels of SEAP secretion in the absence of drug are undetectable, but it is clear that in this system there is at least a 1000-fold enhancement over background levels in the absence of rapamycin. Thus, this system exhibits undetectable background activity and high dynamic range.

Several different configurations for transcription factor fusion proteins were explored. When various numbers of copies of FKBP domains were fused to ZFHD1 and various numbers of copies of FRBs to p65, optimal levels of rapamycin-induced activation occurred when there were multiple FKBP domains fused to ZFHD1 and fewer FRBs fused to p65. The preference for multiple drug-binding domains on the DNA-binding protein may reflect the capacity of these proteins to recruit multiple activation domains and therefore to elicit higher levels of promoter activity. The presence of only 1 drug-binding domain on the activation domain should allow each FKBP on ZFHD to recruit one p65. Any increase in the number of FRBs on p65 would increase the chance that fewer activation domains would be recruited to ZFHD, each one linked by multiple FRB-FKBP interactions.

#### Methods:

HT1080 cells (ATCC CCL-121), derived from a human fibrosarcoma, were grown in MEM supplemented with non-essential amino acids and 10% Fetal Bovine Serum. Cells plated in 24-well dishes (Falcon,  $6 \times 10^4$  cells/well) were transfected using Lipofectamine under conditions recommended by the manufacturer (GIBCO/BRL). A total of 300 ng of the following DNA was transfected into each well: 100 ng ZFHDx12-CMV-SEAP reporter gene, 2.5ng pCGNN-ZFHD1-3FKBP or other DNA binding domain fusion, 5 ng pCGNN-1FRB-p65(361-550) or other activation domain fusion and 192.5 ng pUC118. In cases where the DNA binding domain or activation domain were omitted an equivalent amount of empty pCGNN expression vector was substituted. Following lipofection (for 5 hours) 500  $\mu$ l

medium containing the indicated amounts of rapamycin was added to each well. After 24 hours, medium was removed and assayed for SEAP activity as described (Spencer et al, Science 262:1019-24, 1993) using a Luminescence Spectrometer (Perkin Elmer) at 350 nm excitation and 450 nm emission. Background SEAP activity, measured from mock-transfected

5 cells, was subtracted from each value.

To prepare transiently transfected HT1080 cells for injection into mice (See below), cells in 100 mm dishes ( $2 \times 10^6$  cells/dish) were transfected by calcium phosphate precipitation for 16 hours (Gatz, C., Kaiser, A. & Wendenburg, R., 1991, *Mol. Gen. Genet.* **227**, 229-237) with the following DNAs: 10 mg of ZHWTx12-CMV-hGH, 1 mg pCGNN-ZFHD1-3FKBP, 2  
10 mg pCGNN-1FRB-p65(361-550) and 7 mg pUC118. Transfected cells were rinsed 2 times with phosphate buffered saline (PBS) and given fresh medium for 5 hours. To harvest for injection, cells were removed from the dish in Hepes Buffered Saline Solution containing 10 mM EDTA, washed with PBS/0.1% BSA/0.1% glucose and resuspended in the same at a concentration of  $2 \times 10^7$  cells/ml.

15

#### Plasmids:

Construction of the transcription factor fusion plasmids is described above.

#### pZHWTx12-CMV-SEAP

20 This reporter gene, containing 12 tandem copies of a ZFHD1 binding site (Pomerantz et al., 1995) and a basal promoter from the immediate early gene of human cytomegalovirus (Boshart et. al., 1985) driving expression of a gene encoding secreted alkaline phosphatase (SEAP), was prepared by replacing the NheI-HindIII fragment of pSEAP Promoter (Clontech) with the following NheI-XbaI fragment containing 12 ZFHD binding sites:

25

GCTAGCTAATGATGGGCGCTCGAGTAATGATGGGCGGTCGACTAATGATGGGCGCTCGAGTAATGATGGGCGTC  
TAGCTAATGATGGGCGCTCGAGTAATGATGGGCGGTCGACTAATGATGGGCGCTCGAGTAATGATGGGCGTCTA  
GCTAATGATGGGCGCTCGAGTAATGATGGGCGGTCGACTAATGATGGGCGCTCGAGTAATGATGGGCGTCTAGA  
(the ZFHD1 binding sites are underlined),

30

and the following XbaI-HindIII fragment containing a minimal CMV promoter (-54 to +45):

TCTAGAACCGGAATTCCGGTAGGCGTGTACGGTGGGAGGTCATATAAGCAGAGCTCGTTTAGTGAACCGTCAG  
ATCGCCTGGAGACGCCATCCACGCTGTTTGTACCTCCATAGAAGCTT

35 (the CMV minimal promoter is underlined).

### pZHWTx12-CMV-hGH

Activation of this reporter gene leads to the production of hGH. It was constructed by replacing the HindIII-BamHI (blunted) fragment of pZHWTx12-CMV-SEAP (containing the SEAP coding sequence) with a HindIII (blunted) -EcoRI fragment from p0GH (containing an hGH genomic clone; Selden et al., MCB 6:3171-3179, 1986; the BamHI and EcoRI sites were blunted together).

### pZHWTx12-IL2-SEAP

This reporter gene is identical to pZHWTx12-CMV-SEAP except the XbaI-HindIII fragment containing the minimal CMV promoter was replaced with the following XbaI-HindIII fragment containing a minimal IL2 gene promoter (-72 to +45 with respect to the start site; Siebenlist et al., MCB 6:3042-3049, 1986):

TCTAGAACGCGAATTCAACATTTTGACACCCCCATAATATTTTCCAGAATTAACAGTATAAATTGCATCTCTT  
GTTC AAGAGTTC CCTATCACTCTCTTTAATCACTACTCACAGTAACCTCAACTCCTGCCACAAGCTT  
(the IL2 minimal promoter is underlined).

### pLH

To facilitate the stable integration of a single, or few, copies of reporter gene the following retroviral vector was constructed. pLH (LTR-*hph*), which contains the hygromycin B resistance gene driven by the Moloney murine leukemia virus LTR and a unique internal ClaI site, was constructed as follows: The hph gene was cloned as a HindIII-ClaI fragment from pBabe Hygro (Morganstern and Land, NAR 18:3587-96, 1990) into BamHI-ClaI cut pBabe Bleo (resulting in the loss of the bleo gene; the BamHI and HindIII sites were blunted together).

### pLH-ZHWTx12-IL2-SEAP

To clone a copy of the reporter gene containing 12 tandem copies of the ZFHD1 binding site and a basal promoter from the IL2 gene driving expression of the SEAP gene into the pLH retroviral vector, the MluI-ClaI fragment from pZHWTx12-IL2-SEAP (with ClaI linkers added) was cloned into the ClaI site of pLH. It was oriented such that the directions of transcription from the viral LTR and the internal ZFHD-IL2 promoters were the same.

### pLH-G5-IL2-SEAP

To construct a retroviral vector containing 5 Gal4 sites embedded in a minimal IL2 promoter driving expression of the SEAP gene, a ClaI-BstBI fragment consisting of the following was inserted into the ClaI site of pLH such that the directions of transcription from the viral LTR

and the internal Gal4-IL2 promoters were the same: A ClaI-HindIII fragment containing 5 Gal4 sites (underlined) and regions -324 to -294 (bold) and -72 to +45 of the IL2 gene (italics)

5 5' ATCGATG**TTTTCTGAGTTACTTTTGTATCCCCACCCCCC**CTCGAGCTTGCATGCCTGCAGGTCGGAG  
TACTGTCTCCGAGCGGAGTACTGTCTCCGAGCGGAGTACTGTCTCCGAGCGGAGTACTGTCTCCGAGCGG  
AGTACTGTCTCCGAGCGCAGACTCTAGAGGATCCGAGAACATTTTGACACCCCCATAATATTTTCCAGAATT  
AACAGTATAAATTGCATCTCTTGTTCAGAGTCCCTATCACTCTCTTTAATCACTACTCACAGTAACCTCAAC  
TCCTGCCACAAGCTT,

10 and a HindIII-BstBI fragment containing the SEAP gene coding sequence (Berger et al., Gene 66:1-10, 1988) mutagenized to add the following sequence (containing a BstBI site) immediately after the stop codon:

5'-CCCGTGGTCCCGCGTTGCTTCGAT

## 15 5. Rapamycin-dependent transcriptional activation in stably transfected cells

We conducted the following experiments to confirm that this system exhibits similar properties in stably transfected cells. We generated stable cell lines by sequential transfection of a SEAP target gene and expression vectors for ZFHD1-3FKBP and 1FRB-p65, respectively. A pool of several dozen stable clones resulting from the final transfection exhibited rapamycin-dependent SEAP production. From this pool, we characterized several individual clones, many of which produced high levels of SEAP in response to rapamycin. One such clone produced SEAP at levels approximately forty times higher than the pool and significantly higher than transiently transfected cells. In an attempt to rigorously quantitate background SEAP production and induction ratio in this clone, we performed a second set of assays in which the length of the SEAP assay was increased by a factor of approximately 50 to detect any SEAP activity in untreated cells. Under these conditions, mock transfected cells produced 47 arbitrary fluorescence units, while the transfected clone produced 54 units in the absence of rapamycin and over 90,000 units at 100 nM rapamycin. Thus, in this stable cell line, background gene expression was negligible and the induction ratio (7 units to 90,000 units) was greater than four orders of magnitude.

To simplify the task of stable transfection, we used a bicistronic expression vector that directs the production of both ZFHD1-3FKBP and 1FRB-p65 through the use of an internal ribosome entry sequence (IRES). This expression plasmid was cotransfected, together with a zeocin-resistance marker plasmid, into a cell line carrying a retrovirally-transduced SEAP reporter gene, and a pool of approximately fifty drug-resistant clones was selected and

expanded. This pool of clones also exhibited rapamycin-dependent SEAP production with no detectable background and a very similar dose-response curve to that observed in transiently transfected cells. Our results indicate that rapamycin-responsive gene expression can be readily obtained in both transiently and stably transfected cells. In both cases, regulation is characterized by very low background and high induction ratios.

**Stable cell lines.** Helper-free retroviruses containing the reporter gene or DNA binding domain fusion were generated by transient co-transfection of 293T cells (Pear, W.S., Nolan, G.P., Scott, M.L. & Baltimore D., 1993, Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl. Acad. Sci. USA* **90**, 8392-8396) with a Psi(-) amphotropic packaging vector and the retroviral vectors pLH-ZHWTx12-IL2-SEAP or SMTN-ZFHD1-3FKBP, respectively. To generate a clonal cell line containing the reporter gene stably integrated, HT1080 cells infected with retroviral stock were diluted and selected in the presence of 300 mg/ml Hygromycin B. Individual clones from this and other cell lines described below were screened by transient transfection of the missing components followed by the addition of rapamycin as described above. All 12 clones analyzed were inducible and had little or no basal activity. The most responsive clone, HT1080L, was selected for further study.

HT20-6 cells, which contain the pLH-ZHWTx12-IL2-SEAP reporter gene, ZFHD1-3FKBP DNA binding domain and 1FRB-p65(361-550) activation domain stably integrated, were generated by first infecting HT1080L cells with SMTN-ZFHD1-3FKBP-packaged retrovirus and selecting in medium containing 500 mg/ml G418. A strongly responsive clone, HT1080L3, was then transfected with linearized pCGNN-1FRB-p65(361-550) and pZeoSV (Invitrogen) and selected in medium containing 250 mg/ml Zeocin. Individual clones were first tested for the presence of 1FRB-p65(361-550) by western. Eight positive clones were analyzed by addition of rapamycin. All eight had low basal activity and in six of them, gene expression was induced by at least two orders of magnitude. The clone that gave the strongest response, HT20-6, was selected for further analysis.

HT23 cells were generated by co-transfecting HT1080L cells with linearized pCGNN-1FRB-p65(361-550)-IRES-ZFHD1-3FKBP and pZeoSV and selecting in medium containing 250 mg/ml Zeocin. Approximately 50 clones were pooled for analysis.

For analysis, cells were plated in 96-well dishes ( $1.5 \times 10^4$  cells/well) and 200  $\mu$ l medium containing the indicated amounts of rapamycin (or vehicle) was added to each well. After 18 hours, medium was removed and assayed for SEAP activity. In some cases, medium was diluted before analysis and relative SEAP units obtained multiplied by the fold-dilution. Background SEAP activity, measured from untransfected HT1080 cells, was subtracted from each value.



## 6. Rapamycin-dependent Production of hGH in Mice

***In Vivo Methods: Animals, husbandry, and general procedures.*** Male *nu/nu* mice were obtained from Charles River Laboratories (Wilmington, MA) and allowed to acclimate for five days prior to experimentation. They were housed under sterile conditions, were allowed free access to sterile food and sterile water throughout the entire experiment, and were handled with sterile techniques throughout. No immunocompromised animal demonstrated outward infection or appeared ill as a result of housing, husbandry techniques, or experimental techniques.

To transplant transiently transfected cells into mice,  $2 \times 10^6$  transfected HT1080 cells, were suspended in 100 ml PBS/0.1% BSA/0.1% glucose buffer, and administered into four intramuscular sites (approximately 25 ml per site) on the haunches and flanks of the animals. Control mice received equivalent volume injections of buffer alone.

Rapamycin was formulated for *in vivo* administration by dissolution in equal parts of N,N-dimethylacetamide and a 9:1 (v:v) mixture of polyethylene glycol (average molecular weight of 400) and polyoxyethylene sorbitan monooleate. Concentrations of rapamycin, in the completed formulation, were sufficient to allow for *in vivo* administration of the appropriate dose in a 2.0 ml/kg injection volume. The accuracy of the dosing solutions was confirmed by HPLC analysis prior to intravenous administration into the tail veins. Some control mice, bearing no transfected HT1080 cells, received 10.0 mg/kg rapamycin. In addition, other control mice, bearing transfected cells, received only the rapamycin vehicle. Blood was collected by either anesthetizing or sacrificing mice via CO<sub>2</sub> inhalation.

Anesthetized mice were used to collect 100 ml of blood by cardiac puncture. The mice were revived and allowed to recover for subsequent blood collections. Sacrificed mice were immediately exsanguinated. Blood samples were allowed to clot for 24 hours, at 4°C, and sera were collected following centrifugation at 1000 x g for 15 minutes. Serum hGH was measured by the Boehringer Mannheim non-isotopic sandwich ELISA (Cat No. 1 585 878). The assay had a lower detection limit of 0.0125 ng/ml and a dynamic range that extended to 0.4 ng/ml. Recommended assay instructions were followed. Absorbance was read at 405 nm with a 490 nm reference wavelength on a Molecular Devices microtiter plate reader. The antibody reagents in the ELISA demonstrate no cross reactivity with endogenous, murine hGH in diluent sera or native samples.

**hGH expression *In Vivo*.** For the assessment of dose-dependent rapamycin-induced stimulation of hGH expression, rapamycin was administered to mice approximately one hour following injection of HT1080 cells. Rapamycin doses were either 0.01, 0.03, 0.1, 0.3, 1.0,

3.0, or 10.0 mg/kg. Seventeen hours following rapamycin administration, the mice were sacrificed for blood collection.

To address the time course of *in vivo* hGH expression, mice received 10.0 mg/kg of rapamycin one hour following injection of the cells. Mice were sacrificed at 4, 8, 17, 24, and

5 42 hours following rapamycin administration.

The ability of rapamycin to induce sustained expression of hGH from transplanted HT1080 cells was tested by repeatedly administering rapamycin. Mice were administered transfected HT1080 cells as described above. Approximately one hour following injection of the cells, mice received the first of five intravenous 10.0 mg/kg doses of rapamycin. The four  
10 remaining doses were given under anesthesia, immediately subsequent to blood collection, at 16, 32, 48, and 64 hours. Additional blood collections were also performed at 72, 80, 88, and 96 hours following the first rapamycin dose. Control mice were administered cells, but received only vehicle at the various times of administration of rapamycin. Experimental  
15 animals and their control counterparts were each assigned to one of two groups. Each of the two experimental groups and two control groups received identical drug or vehicle treatments, respectively. The groups differed in that blood collection times were alternated between the two groups to reduce the frequency of blood collection for each animal.

## Results

20 Rapamycin elicited dose-responsive production of hGH in these animals (Fig. 1). hGH concentrations in the rapamycin-treated animals compared favorably with normal circulating levels in humans (0.2–0.3 ng/ml). No plateau in hGH production was observed in these experiments, suggesting that the maximal capacity of the transfected cells for hGH production was not reached. Control animals—those that received transfected cells but no rapamycin  
25 and those that received rapamycin but no cells—exhibited no detectable serum hGH. Thus, the production of hGH in these animals was absolutely dependent upon the presence of both engineered cells and rapamycin.

The presence of significant levels of hGH in the serum 17 hours after rapamycin administration was noteworthy, because hGH is cleared from the circulation with a half-life of  
30 less than four minutes in these animals. This observation suggested that the engineered cells continued to secrete hGH for many hours following rapamycin treatment. To examine the kinetics of rapamycin control of hGH production, we treated animals with a single dose of rapamycin and then measured hGH levels at different times thereafter. Serum hGH was observed within four hours of rapamycin treatment, peaked at eight hours (at over one  
35 hundred times the sensitivity limit of the hGH ELISA), and remained detectable 42 hours after treatment. hGH concentration decayed from its peak with a half-life of approximately 11 hours. This half-life is several hundredfold longer than the half-life of hGH itself and

approximately twice the half-life of rapamycin (4.6 hr) in these animals. The slower decay of serum hGH relative to rapamycin could reflect the presence of higher tissue concentrations of rapamycin in the vicinity of the implanted cells. Alternatively, persistence of hGH production from the engineered cells may be enhanced by the stability of hGH mRNA.

5 Interestingly, administration of a second dose of rapamycin to these animals at 42 hr resulted in a second peak of serum hGH, which decayed with similar kinetics indicating that the engineered cells retained the ability to respond to rapamycin for at least two days. Therefore, to ascertain the ability of this system to elevate and maintain circulating hGH concentrations, we performed an experiment in which animals received multiple doses of rapamycin at 16-  
10 hour intervals. This interval corresponds to the time required for hGH levels to peak and then decline approximately half-way. According to this regimen, rapamycin concentration is predicted to approach a steady-state trough concentration of 1.7 µg/ml after two doses. hGH levels should also approach a steady state trough concentration following the second dose. Indeed, treated animals held relatively stable levels of circulating hGH in response to  
15 repeated doses of rapamycin. After the final dose, hGH levels remained constant for 16 hours and then declined with a similar half-life as rapamycin (6.8 hours for hGH versus 4.6 hours for rapamycin). These data suggest that upon multiple dosing, circulating rapamycin imparts tight control over the secretion of hGH from transfected cells *in vivo*. In particular, it is apparent that protein production is rapidly terminated upon withdrawal of drug.

20

## Discussion

These experiments demonstrate that the transcription factor component modules function appropriately with corresponding target gene constructs in cell culture and in whole animals in a regulatable system.

25

## B. Hybrid transcription factors containing such modular components work well in constitutive expression

### Plasmids

30

#### pCGNNZFHD1

An expression vector for directing the expression of ZFHD1 coding sequence in mammalian cells was prepared as follows. Zif268 sequences were amplified from a cDNA clone by PCR using primers 5'Xba/Zif and 3'Zif+G. Oct1 homeodomain sequences were amplified from a  
35 cDNA clone by PCR using primers 5'Not Oct HD and Spe/Bam 3'Oct. The Zif268 PCR fragment was cut with XbaI and NotI. The Oct1 PCR fragment was cut with NotI and BamHI. Both fragments were ligated in a 3-way ligation between the XbaI and BamHI sites of

pCGNN (Attar and Gilman, 1992) to make pCGNNZFHD1 in which the cDNA insert is under the transcriptional control of human CMV promoter and enhancer sequences and is linked to the nuclear localization sequence from SV40 T antigen. The plasmid pCGNN also contains a gene for ampicillin resistance which can serve as a selectable marker.

5

#### pCGNNZFHD1-p65

An expression vector for directing the expression in mammalian cells of a chimeric transcription factor containing the composite DNA-binding domain, ZFHD1, and a transcription activation domain from p65 (human) was prepared as follows. The sequence encoding the C-terminal region of p65 containing the activation domain (amino acid residues 450-550) was amplified from pCGN-p65 using primers p65 5' Xba and p65 3' Spe/Bam. The PCR fragment was digested with Xba1 and BamH1 and ligated between the the Spe1 and BamH1 sites of pCGNN ZFHD1 to form pCGNN ZFHD-p65AD.

15

The P65 transcription activation sequence contains the following linear sequence:

CTGGGGGCGCTTGCTTGGCAACAGCACAGACCCAGCTGTGTTCACAGACCTGGCATCCGTCGACAACTCCGAGTT  
TCAGCAGCTGCTGAACAGGGCATACTGTGGCCCCCACACAACCTGAGCCCATGCTGATGGAGTACCCCTGAGG  
CTATAACTCGCCTAGTGACAGGGGCCCCAGAGGCCCCCGACCCAGCTCCTGCTCCACTGGGGGCCCCGGGGCTC  
CCCAATGGCCTCCTTTTCAGGAGATGAAGACTTCTCCTCCATTGCGGACATGGACTTCTCAGCCCTGCTGAGTCA  
GATCAGCTCC

#### pCGNNZFHD1-FKBPx3

An expression vector for directing the expression of ZFHD1 linked to three tandem repeats of human FKBP was prepared as follows. Three tandem repeats of human FKBP were isolated as an XbaI-BamHI fragment from pCGNNF3 and ligated between the Spe1 and BamHI sites of pCGNNZFHD1 to make pCGNNZFHD1-FKBPx3 (ATCC Accession No. 97399).

#### 30 pZHWTx8SVSEAP

A reporter gene construct containing eight tandem copies of a ZFHD1 binding site (Pomerantz *et al.*, 1995) and a gene encoding secreted alkaline phosphatase (SEAP) was prepared by ligating the tandem ZFHD1 binding sites between the Nhe1 and BglII sites of pSEAP-Promoter Vector (Clontech) to form pZHWTx8SVSEAP. The ZHWTx8SEAP reporter contains two copies of the following sequence in tandem:

35

CTAGCTAATGATGGGCGCTCGAGTAATGATGGGCGCTCGACTAATGATGGGCGCTCGAGTAATGATGGGCGT

The ZFHD1 binding sites are underlined.

#### 5 pCGNN F1 and F2

One or two copies of FKBP12 were amplified from pNF3VE using primers FKBP 5' Xba and FKBP 3' Spe/ Bam. The PCR fragments were digested with Xba1 and BamH1 and ligated between the Xba1 and BamH1 sites of pCGNN vector to make pCGNN F1 or pPCGNN F2. pCGNNZFHD1-FKBPx3 can serve as an alternate source of the FKBP cDNA.

10

#### pCGNN F3

A fragment containing two tandem copies of FKBP was excised from pCGNN F2 by digesting with Xba1 and BamH1. This fragment was ligated between the Spe1 and BamH1 sites of pCGNN F1.

15

#### pCGNN F3VP16

The C-terminal region of the Herpes Simplex Virus protein, VP16 (AA 418-490) containing the activation domain was amplified from pCG-Gal4-VP16 using primers VP16 5' Xba and VP16 3' Spe/Bam. The PCR fragment was digested with Xba1 and BamH1 and ligated between the Spe1 and BamH1 sites of pCGNN F3 plasmid.

20

#### pCGNN F3p65

The Xba1 and BamH1 fragment of p65 containing the activation domain was prepared as described above. This fragment was ligated between the Spe1 and BamH1 sites of pCGNN F3.

25

#### Primers

|           |                                     |
|-----------|-------------------------------------|
| 5'Xba/Zif | 5'ATGCTCTAGAGAACGCCCATATGCTTGCCCT   |
| 3'Zif+G   | 5'ATGCGCGGCCGCGCCTGTGTGGGTGCGGATGTG |

30

|               |                                                   |
|---------------|---------------------------------------------------|
| 5'Not OctHD   | 5'ATGCGCGGCCGCGCAGGAGGAAGAAACGCACCAGC             |
| Spe/Bam 3'Oct | 5'GCATGGATCCGATTCAACTAGTGTGATTCTTTTTCTTTCTGGCGGCG |

FKBP 5'Xba

5'TCAGTCTAGAGGAGTGCAGGTGGAACCAT

35 FKBP 3' Spe/Bam

5'TCAGGGATCCTCAATAACTAGTTTCCAGTTTTAGAAGCTC

VP16 5' Xba 5'ACTGTCTAGAGTCAGCCTGGGGGACGAG  
 VP16 3' Spe/Bam 5'GCATGGATCCGATTCAACTAGTCCCACCGTACTCGTCAATTCC

P65 5' Xba 5'ATGCTCTAGACTGGGGGCCTTGCTTGGCAAC  
 5 p65 3' Spe/Bam 5'GCATGGATCCGCTCAACTAGTGGAGCTGATCTGACTCAG

## References

1. Attar, R.M., and M.Z. Gilman 1992. *Mol. Cell. Biol.* 12:2432-2443
2. Ausubel, F.M. *et al.*, Eds., 1994. CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Wiley, NY)
- 10 3. Pomerantz, J.L., *et al.* 1995. *Science*. 267:93-96.
4. Spencer, D.M., *et al.* 1993. *Science*. 262:1019-1024.

## II. Evaluation of representative illustrative chimeric transcription factors

15

### Constructs

Constructs encoding the following GAL-4-based chimeric transcription factors, among others, were prepared and tested in human cell lines containing stably integrated SEAP reporter constructs containing GAL4 or ZFHD1 recognition sequences, as appropriate:

20

|                 |                             |
|-----------------|-----------------------------|
| chimeric factor | <u>data shown in Figure</u> |
|-----------------|-----------------------------|

Fig. 2

G-K

G-KK

G-KKK

25

G-KKKK

G-KKKKK

G-KKKKKK

Fig. 3

G-(V8x2)

30

G-(V8x2)2

G-(V8x2)3

G-(V8x2)4

G-(V8x2)5

G-(V8x2)6

35

(continued —>)

(continued from previous page)

| chimeric factor        | data shown in Figure |
|------------------------|----------------------|
| G-D                    | Fig. 4               |
| 5 G-DD                 |                      |
| G-DDD                  |                      |
| G-DDDD                 |                      |
| G-DDDDD                |                      |
| G-DDDDDD               |                      |
| 10                     |                      |
| Z-VP16                 | Fig. 5               |
| Z-k                    |                      |
| Z-kkk                  |                      |
| Z-K                    |                      |
| 15 Z-KKK               |                      |
| G-KKK-(V8x2)4          | Fig. 6               |
| G-KKK-DDDDD            |                      |
| G-(V8x2)4-DDDDD        |                      |
| 20 G-KKK-(V8x2)4-DDDDD |                      |
| G-K                    | Fig. 7               |
| G-KKK                  |                      |
| G-HSF-HSF              |                      |
| 25 G-HSF-HSF-HSF-HSF   |                      |
| G-K-HSF-HSF-HSF-HSF    |                      |
| G-KKK-HSF-HSF-HSF-HSF  |                      |

- 
- 30 abbreviations: G = GAL4 residues 1-94  
 K = p65(361-550) = "N361" in Fig. 6  
 k = p65(450-550) = "N450" in Fig. 6  
 V8x2 = tandem repeat of VP16 V8 sequence with an intervening  
 SerArg resulting from ligation; (V8x2)4="8V8" in Fig 6  
 D = VP16 C terminal SRDFDLMLG containing an initial SerArg  
 35 resulting from ligation = "Vc" in Fig 6  
 Z = ZFHD1 ("ZH" in Fig 5)  
 HSF = 14 mer (see table below)

**Plasmid constructions:** PCG-Gal4 vector containing Gal4 DNA binding domain coding sequences between amino acids 1-94 was digested with Xba1 and BamH1. The p65 activation domain sequences between amino acids 361-550 was generated by PCR using the following oligonucleotides:

5'-atgctctagagatgagttcccacatggtg-3'

and

5'-gcatggatccgctcaactagtggagctgactgactcag-3'.

This fragment was digested with Xba1 and BamH1 and cloned into PCG-Gal4 vector to make PCG-Gal4-p65 (361-550), here after will be referred as PCG-GK. To make PCG-GK2 plasmid, the p65 activation domain containing PCR fragment described above was digested with Xba1 and BamH1 and cloned into Spe1 and BamH1 digested PCG-GK vector. PCG-GK3, 4, 5,6 were all generated following the same approach.

Plasmid PCG-Gal 4 plasmids containing reiterated copies of V8 domain were generated by the following method. The oligonucleotides 5'-ctagagactcgacttgacatgct-3'; 5'-agtccccagcatgtccaagtcgaagtct-3'; 5'-gggggacttcgacttgacatgctgactagttgag-3' and 5'-gactcctcaactagtcagcatgtccaagtcga-3' were phosphorylated and the first and last pair of oligos were annealed separately. Together these oligonucleotides make two tandem V8 coding sequences. These annealed oligos were then ligated into Xba1 and BamH1 digested PCG-Gal4 vector. The resulting vector, PCG-GV2 containing two copies of V8 sequences was digested with Spe1 and BamH1. V8x2 oligos made as described above was cloned into this vector to make PCG-GV4. Same approach was taken to generate PCG-GV6, 8, 10 and 12 plasmids.

PCG-Gal4 plasmids containing reiterated copies of VP16 C-terminus, hereafter referred as D activation domain were constructed as follows. The VP16 C-terminus region was PCR amplified using the following primers: 5'-atgctctagagacggggattccccggggccg-3' and 5'-gcatggatcctcattaactagtcccacgctactcgtaattcc-3'. The PCR fragments were digested with Xba1 and BamH1 and cloned into PCG-Gal4 vector previously digested with Xba1 and BamH1. The resulting plasmid was designated as PCG-GD. To make PCG-GD2, PCG-GD was digested with Spe1 and BamH1 and ligated with Xba1 and BamH1 digested D fragment described above. PCG-GD3,4,5 and 6 were constructed using the same approach.



- Plasmids PCG-GK3V8 and PCG-GK3D5 were made by digesting PCG-GV8 and PCG-D5 plasmids with Xba1 and BamH1 and cloning the fragments containing V8 and D5 sequences respectively into PCG-GK3 digested with Spe1 and BamH1. Similarly, Xba1/BamH1
- 5 fragment from PCG-GD5 containing D5 sequences was cloned into Spe1/BamH1 digested PCG-GV8 plasmid to construct PCG-V8D5 plasmid. The V8D5 fragment was excised from this plasmid by digesting it with Xba1 and BamH1 and the fragment was cloned into Spe1/BamH1 digested PCG-K3 to make PCG-K3V8D5 plasmid.
- 10 PCGNN-ZFHD-p65(450-550) and PCGNN-ZFHD-p65(361-550) are described above. PCGNN-p65(450-550)x3 and PCGNN-ZFHD-p65(361-550) were made as follows: PCG-Gal4-p65(450-550)x3 and PCG-Gal4-p65(361-550) were digested with Xba1 and BamH1 and the p65(450-550)x3 and p65(361-550) were excised. These fragments were cloned into Spe1/BamH1 digested PCGNN-ZFHD to generate PCGNN-ZFHD-p65(450-550) and
- 15 PCGNN-ZFHD-p65(361-550).

PCG-Gal4-HSFX2 containing two copies of HSF14 activation domain was made by phosphorylating and ligating the following oligonucleotides to Xba1 and BamH1 digested PCG-Gal4 plasmid:

- 20 5'-ctagagacaccagtgccctgctggacctgttcagcccctcg-3';  
 5'-ggtcaccgaggggctgaacaggtccagcagggcactggtgtct-3';  
 5'-gtgaccgtgcccgacatgagcctgcctgaccttgacagcag-3' and  
 5'-gtgaccgtgcccgacatgagcctgcctgaccttgacagcag-3'.

- 25 Two additional copies of HSF activation domain were added to Spe1/BamH1 digested PCG-Gal4-HSFX2 plasmid by the same method to generate PCG-Gal4-HSFX4 plasmid. A fragment containing four copies of HSF14 activation domain was excised from PCG-Gal4-HSFX4 by Xba1 and BamH1 digestion. The resulting fragment was cloned into Spe1 and
- 30 BamH1 digested PCG-Gal4KX1 and PCG-Gal4KX3 to make PCG-Gal4-K+HSFX4 or PCG-Gal4-K3+HSFX4 plasmids.

### reporter cell lines

- Human 1080 cells were engineered by the stable introduction of a secreted alkaline phosphatase (SEAP) target gene construct. The target gene construct contained a gene
- 35 encoding SEAP operably linked to a transcription control sequence containing five copies of a DNA recognitions sequence for GAL4 and a minimal IL-2 promoter. The resultant cells may be

used in experiments such as described in Example 3 in which the cells are further transfected with DNA constructs encoding various transcription factors containing one or more DNA binding domains recognized by the target gene construct.

5 **plasmid constructions: pLH-G5-IL2-SEAP** (as previously described)

**cell culture:** HT1080 cells (ATCC CCL-121), derived from a human fibrosarcoma, were grown in MEM supplemented with non-essential amino acids and 10% Fetal Bovine Serum. Helper-free retroviruses containing the 5xGAL4-IL2-SEAP reporter gene were generated by  
10 transient co-transfection of 293T cells (Pear, W.S., Nolan, G.P., Scott, M.L. & Baltimore, D. *USA* **90**, 8392-8396 (1993) with a Psi(-) amphotropic packaging vector and the retroviral vector pLH-5xGAL4-IL2-SEAP. To generate a clonal cell line containing the SEAP reporter gene stably integrated, HT1080 cells infected with retroviral stock were diluted and selected  
15 in the presence of 300 mg/ml Hygromycin B. Individual clones were screened for the presence of integrated reporter gene by transient transfection of a plasmid encoding a chimeric transcription factor containing a GAL4 DNA binding domain. The most responsive clone, HT1080B, was used for subsequent analysis.

20 **Analysis of chimeric transcription factors**

Transfection: HT1080 B cells were grown in MEM supplemented with 10% Bovine Calf Serum. Approximately 2X10<sup>5</sup> cells/well in a 12 well plate were transiently transfected by Lipofectamine procedure as recommended by GIBCO, BRL. The DNA:Lipofectamine ratio  
25 used correspond to 1:6. Cells in each well recieved indicated amounts of effector plasmids and total DNA concentration in each well was adjusted to 1.25 ug with PUC118 DNA. Following transfection, 1ml fresh media was added to each well. After 24 hrs, 100ul of the media was assayed for SEAP activity as described.

### Representative results:

| chimeric transcription factor              | number of<br>activation<br>domains | transcription<br>activation<br>(IL2 promoter) |
|--------------------------------------------|------------------------------------|-----------------------------------------------|
| GAL4-p65(361-550)                          | 1 to 6                             | ++++                                          |
| GAL4-p65(450-550)                          | 1 to 6                             | +++                                           |
| GAL4-p65(361-450)                          | 1 to 6                             | --                                            |
| GAL4-K13 (SRDFADMDFDALL, derived from p65) | 1 to 6                             | +++                                           |
| GAL4-Oct2 Q domain (aa95-160)              | 1 to 6                             | --                                            |
| GAL4-Oct2 P domain (aa438-479)             | 1 to 6                             | --                                            |
| GAL4-HSF (aa 409-444)                      | 1 to 4                             | +++                                           |
| GAL4-HSF14 (DLDSSLASIQELLS)                | 1 to 4                             | ++                                            |
| GAL4-EWS11 (SRSYGQQGSGS)                   | 1 to 8                             | --                                            |
| GAL4-V8x2 (DFDLMLGDFDLMLGSR)               | 1 to 12                            | ++                                            |
| GAL4-D (VP16 aa 459-490)                   | 1 to 6                             | +++                                           |
| GAL4-VP16 (VP16 aa 411-490)                | 1 to 4                             | ++                                            |

### III. Illustrative chimeric transcription factor for allostery-based systems

#### p65/GAL4/PR-LBD

An expression vector for directing the expression of an RU486 dependent transcription factor consisting of a progesterone receptor ligand binding domain, a GAL4 DNA binding domain and a p65 transcription activation domain can be prepared as follows. Primers 5'-p65-BglII and 3'-p65-BamHI can be used to amplify amino acids 361 to 550 of p65 from plasmid pCG-p65 (Rivera et al., Nature Medicine 2(9):1028-1032, 1996). The resulting fragment, which will have 5' BglII and 3' BamHI sites, can then be inserted into the BglII site of plasmid pGL (Wang et al, PNAS USA 91:8180-8184, 1994), which contains a truncated human progesterone receptor sequence (amino acids 640-891) and a GAL4 DNA binding domain sequence (amino acids 1-94). Up to 2 nucleotides may be added so that the p65 sequence is in frame downstream of the ATG and upstream of the GAL4 coding region.

5'-p65-BglII: agatctXGATGAGTTTCCCACCATG  
3'-p65-BamHI: ggatccXGGAGCTGATCTGACTCAG

where X is 0, 1 or 2 nucleotides that may be required to create in-frame fusions.

### **ZFHD1/p65/PR-LBD**

An additional RU 486 dependent transcription factor can be prepared using the composite DNA binding domain ZFHD1 (Rivera et al., supra, and US 08/366,083.) Primers 5'-PR-LBD and 3'-PR-LBD can be used to amplify amino acids 640-891 of hPRB891 from plasmid pT7bhPRB-891 (Vegeto et al, Cell 69:703-713, 1992). The resulting fragment, which will have 5' XbaI and 3' SpeI sites, can be inserted into the SpeI site of pCGNN-ZFHD1-p65. This will place the PR-LBD in-frame and at the carboxy terminus of p65.

5'-PR-LBD: 5'-tctagaAAAAAGTTCAATAAAGTCAG

3'-PR-LBD: 5'-actagtGCAGTACAGATGAAGTTG

### **rtTA/p65**

A tetracycline inducible transcription factor containing the p65 activation domain can be constructed using pUHD17-1 (described in US 5,654,168) as follows. Digest pUHD17-1 with AflIII. Remove the protruding 5' end with mung bean nuclease and ligate the synthetic oligonucleotide 5'-CactagtTAACTAAGTAA. The resulting plasmid, rTetR-SpeI contains a SpeI cleavage site at the very end of the rTetR gene.

Use primers 5'-p65-XbaI and 3'-p65-SpeI to amplify amino acids 361 to 550 of p65 from plasmid pCG-p65 (Rivera et al, supra) and clone this fragment, in frame, into the SpeI site of rTetR-SpeI.

5'-p65-XbaI: 5'-tctagaGATGAGTTTCCCACCATG

3'-p65-SpeI: 5'-actagtGGAGCTGATCTGACTCAG

## **IV. Regulated transcription mediated by the chimeric transcription factor S3H**

The S3H activation domain was used to induce transcription of a SEAP target gene in response to rapamycin. S3H (p65(281-551)-HSF(406-530))- and S (p65(361-551)) - containing activation domain fusions were stably integrated into HT1080L cells (which already carry a stably integrated target SEAP gene) as part of the transcription factor vectors

pR5N<sub>2</sub>EN-RH<sub>1</sub>S3H/Z1F3/Neo and pR5N<sub>2</sub>EN-RH<sub>1</sub>S/Z1F3/Neo, respectively. These vectors express an activation domain fusion and a DNA binding domain fusion, separated by an IRES element, from the RSV promoter/enhancer. The activation domain fusion contains a *c-myc* nuclear localization signal (N<sub>2</sub>) fused to the T2098L mutant FRB domain (RH<sub>1</sub>) fused to either the S3H or S activation domains. The DNA binding domain consists of an HA epitope tag (E), a nuclear localization signal from SV40 (N), the ZFHD1 DNA binding domain and 3 FKBP domains. As seen in figure 8, expression of SEAP was measured as a function of increasing doses of rapamycin. Since the RSV enhancer works relatively poorly in HT1080 cells, levels of the transcription factor fusions are relatively low (ie. compared to their levels when expressed from the CMV enhancer/promoter). When expressed at low levels, the S activation domain is not potent enough to support rapamycin induced target gene expression. However, equivalent expression of the S3H activation domain fusion, gives rise to high level activated transcription.